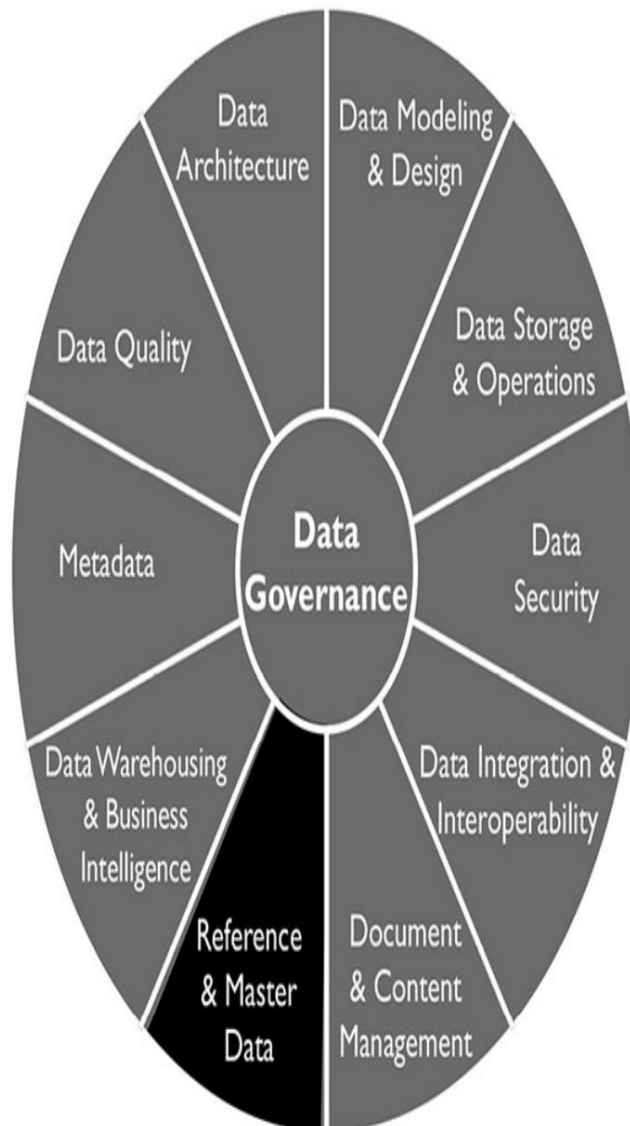


CAPÍTULO 10

Dados de referência e mestre



DAMA-DMBOK2 Data Management Framework

Copyright © 2017 by DAMA International

1. Introdução

Em qualquer organização, certos dados são necessários em todas as áreas de negócios,

processos e sistemas. A organização geral e seus clientes se beneficiam se esses dados forem compartilhados e todas as unidades de negócios puderem acessar as mesmas listas de clientes, códigos de localização geográfica, listas de unidades de negócios, opções de entrega, listas de peças, códigos de centro de custo contábil, códigos de impostos governamentais e outros dados usados para administrar o negócio. Pessoas que usam dados geralmente assumem que existe um nível de consistência em toda a organização, até que vejam dados díspares.

Na maioria das organizações, os sistemas e dados evoluem mais organicamente do que os profissionais de gerenciamento de dados gostariam. Particularmente em grandes organizações, vários projetos e iniciativas, fusões e aquisições e outras atividades comerciais resultam em múltiplos 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 riscos. Ambos podem ser reduzidos por meio do gerenciamento de Master Data e Reference Data.

Figura 75 Diagrama de Contexto: Dados de referência e mestre

Reference and Master Data

Definition: Managing shared data to meet organizational goals, reduce risks associated with data redundancy, ensure higher quality, and reduce the costs of data integration.

Goals:

1. Enable sharing of information assets across business domains and applications within an organization.
2. Provide authoritative source of reconciled and quality-assessed master and reference data.
3. Lower cost and complexity through use of standards, common data models, and integration patterns.

Business
Drivers
↓

Inputs:

- Business Drivers
- Cross Functional Requirements
- Industry Standards
- Data Glossary
- Purchased Data and/or Open Data and Code Sets
- Business Rules

Activities:

1. **Identify Drivers and Requirements (P)**
 1. Validate Data Definitions (C)
2. **Evaluate and Assess Data Sources (P)**
3. **Define Architectural Approach (D)**
4. **Model Data (D)**
5. **Define Stewardship and Maintenance Processes (C)**
6. **Establish Governance Policies (C)**
7. **Implement Data Sharing/Integration Services (D,O)**
 1. Acquire Data Sources for Sharing
 2. Publish Reference and Master Data

Deliverables:

- Master and Reference Data Requirements
- Data Models and Integration Patterns
- Reliable Reference and Master Data
- Reusable Data Services

Suppliers:

- Subject Matter Experts
- Data Stewards
- Application Developers
- Data Providers
- Business Analysts
- Infrastructure Systems Analysts

Participants:

- Data Analysts
- Data Modelers
- Data Stewards
- Data Integrators
- Data Architects
- Data Quality Analysts

Consumers:

- Master Data Analysts
- Data Integrators
- Data Architects
- Application Users
- Application Developers
- Solution Architects

Technical
Drivers
↑

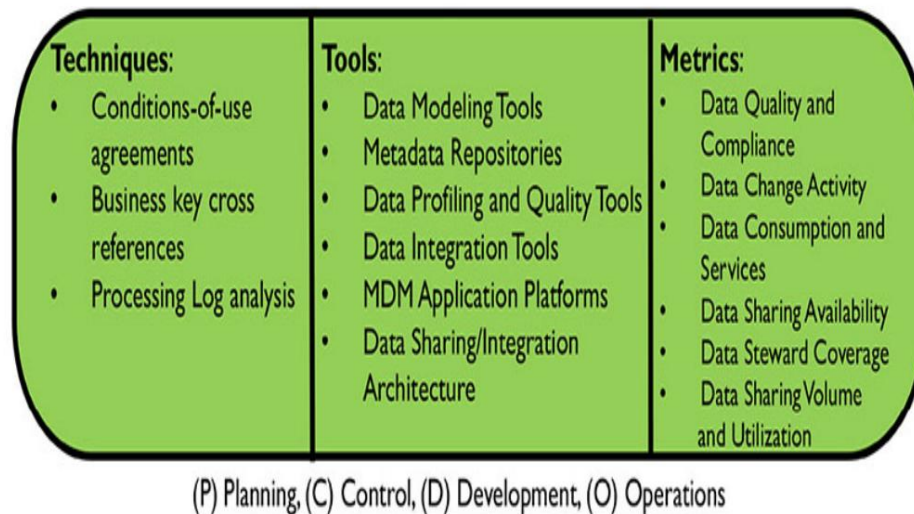


Figure 75 Context Diagram: Reference and Master Data

1.1 Drivers de negócios

Os motivadores mais comuns para iniciar um programa de Gerenciamento de Dados Mestres são:

- **Atendendo aos requisitos de dados organizacionais:** Várias áreas dentro de uma organização precisam acessar os mesmos conjuntos de dados, com a confiança de que os conjuntos de dados são completos, atuais e consistentes. Os Dados Mestres geralmente formam a base desses conjuntos de dados (por exemplo, determinar se uma análise inclui todos os clientes depende de ter uma definição de cliente aplicada de forma consistente).
- **Gerenciamento da qualidade dos dados:** inconsistências de dados, problemas de qualidade e lacunas levam a decisões incorretas ou oportunidades perdidas; o Master Data Management reduz esses riscos ao permitir uma representação consistente das entidades críticas para a organização.
- **Gerenciando os custos de integração de dados:** O custo de integração de novas fontes de dados em um ambiente já complexo é maior na ausência de Dados Mestres, o que reduz a variação na forma como entidades críticas são definidas e identificadas.

- **Redução de riscos:** o Master Data pode permitir a simplificação da arquitetura de compartilhamento de dados para reduzir custos e riscos associados a um ambiente complexo.

Os drivers para gerenciar Dados de Referência são semelhantes. Dados de Referência gerenciados centralmente permitem que as organizações:

- Atender aos requisitos de dados para várias iniciativas e reduzir os riscos e custos da integração de dados por meio do uso de Dados de Referência consistentes
- Gerenciar a qualidade dos Dados de Referência

Enquanto iniciativas organizacionais orientadas por dados focam em dados transacionais (aumento de vendas ou participação de mercado, redução de custos, demonstração de conformidade), a capacidade de alavancar tais dados transacionais é altamente dependente da disponibilidade e qualidade de Dados Mestres e de Referência. Melhorar a disponibilidade e qualidade de Dados Mestres e de Referência tem um impacto dramático na qualidade geral dos dados e na confiança empresarial nos dados. Esses processos têm benefícios adicionais para uma organização, incluindo simplificação do cenário de TI, eficiência e produtividade aprimoradas e, com isso, o potencial de melhorar a experiência do cliente.

1.2 Objetivos e princípios Os objetivos

de um programa de gerenciamento de dados mestres e de referência incluem:

- Garantir que a organização tenha Dados Mestres e de Referência completos, consistentes, atuais e confiáveis em todos os processos organizacionais
- Permitir que dados mestres e de referência sejam compartilhados entre funções e aplicativos empresariais

- Reduzir o custo e a complexidade do uso e integração de dados por meio de padrões, modelos de dados comuns e padrões de integração

O gerenciamento de dados mestres e de referência segue estes princípios orientadores:

- **Dados compartilhados:** os dados de referência e mestres devem ser gerenciados para que sejam compartilháveis em toda a organização.
- **Propriedade:** Os dados de referência e mestres pertencem à organização, não a um aplicativo ou departamento específico. Por serem amplamente compartilhados, eles exigem um alto nível de administração.
- **Qualidade:** O gerenciamento de dados mestres e de referência exige monitoramento e governança contínuos da qualidade dos dados.
- **Administração:** Os administradores de dados comerciais são responsáveis por controlar e garantir a qualidade dos dados de referência.
- **Mudança controlada:**
 - Em um determinado momento, os valores dos Dados Mestres devem representar o melhor entendimento da organização sobre o que é preciso e atual. Regras de correspondência que alteram valores devem ser aplicadas com cautela e supervisão. Qualquer identificador mesclado ou dividido deve ser reversível.
 - Alterações nos valores dos Dados de Referência devem seguir um processo definido; as alterações devem ser aprovadas e comunicadas antes de serem implementadas.
- **Autoridade:** Os valores dos Dados Mestres devem ser replicados somente do sistema de registro. Um sistema de referência pode ser necessário para permitir o compartilhamento de Dados Mestres em um

organização.

1.3 Conceitos Essenciais

1.3.1 Diferenças entre dados mestre e de referência Diferentes

tipos de dados desempenham diferentes funções dentro de uma organização. Eles também têm diferentes requisitos de gerenciamento. Uma distinção é frequentemente feita entre Dados de Transação e Dados Mestres, bem como entre Dados Mestres e Dados de Referência. Malcolm Chisholm propôs uma taxonomia de dados de seis camadas que inclui Metadados, Dados de Referência, dados de estrutura empresarial, dados de estrutura de transação, dados de atividade de transação e dados de auditoria de transação (Chisholm, 2008; Talburt e Zhou, 2015). Dentro dessa taxonomia, ele define Dados Mestres como uma agregação de Dados de Referência, dados de estrutura empresarial e dados de estrutura de transação:

- **Dados de referência**, por exemplo, tabelas de código e descrição, são dados usados exclusivamente para caracterizar outros dados em uma organização ou exclusivamente para relacionar dados em um banco de dados a informações além dos limites da organização.
- **Dados de estrutura empresarial**, por exemplo, um plano de contas, permitem relatar atividades comerciais por responsabilidade empresarial.
- **Dados de estrutura de transação**, por exemplo, identificadores de clientes, descrevem os itens que devem estar presentes para que uma transação ocorra: produtos, clientes, fornecedores.

A definição de Chisholm distingue Master Data de dados de atividade de transação que registram detalhes sobre transações e de dados de auditoria de transação que descrevem o estado das transações, bem como de Metadados, que descrevem outros dados (Chisholm, 2008). A este respeito, a definição de Chisholm é semelhante à definição do Dicionário DAMA: *Master Data* é “os dados que fornecem a

contexto para dados de atividade empresarial na forma de conceitos comuns e abstratos que se relacionam com a atividade. Inclui os detalhes (definições e identificadores) de objetos internos e externos envolvidos em transações empresariais, como clientes, produtos, funcionários, fornecedores e domínios controlados (valores de código)” (DAMA, 2009).

Muitas pessoas entendem que Master Data inclui dados de estrutura de transação e dados de estrutura empresarial. A definição de Master Data de David Loshin se alinha amplamente com esses tipos. Ele descreve objetos Master Data como objetos de negócios principais usados em diferentes aplicativos em uma organização, junto com seus Metadados, atributos, definições, funções, conexões e taxonomias associados. Objetos Master Data representam aquelas "coisas" que mais importam para uma organização – aquelas que são registradas em transações, relatadas, medidas, analisadas (Loshin, 2008).

O Master Data requer a identificação e/ou o desenvolvimento de uma versão confiável da verdade para cada instância de entidades conceituais, como produto, local, conta, pessoa ou organização, e a manutenção da atualidade dessa versão. O principal desafio com o Master Data é a resolução de entidades (também chamada de gerenciamento de identidade), o processo de discernir e gerenciar associações entre dados de diferentes sistemas e processos. As instâncias de entidades representadas por linhas do Master Data serão representadas de forma diferente entre os sistemas. O Master Data Management trabalha para resolver essas diferenças a fim de identificar consistentemente instâncias de entidades individuais (ou seja, clientes específicos, produtos, etc.) em diferentes contextos. Esse processo também deve ser gerenciado ao longo do tempo, para que os identificadores dessas instâncias de entidades do Master Data permaneçam consistentes.⁵⁶ Os Dados de Referência e os ---

Master Data compartilham propósitos conceitualmente semelhantes. Ambos fornecem contexto crítico para a criação e uso de dados transacionais. (Dados de referência também fornecem contexto para Master Data.) Eles permitem que os dados sejam compreendidos de forma significativa. Mais importante, ambos são recursos compartilhados que devem ser gerenciados no nível empresarial. Ter múltiplas instâncias dos mesmos Dados de Referência é ineficiente e inevitavelmente leva à inconsistência entre eles. A inconsistência leva à ambiguidade, e a ambiguidade introduz risco para uma organização.

Um programa bem-sucedido de gerenciamento de dados de referência ou dados mestres envolve toda a gama de funções de gerenciamento de dados (governança de dados, qualidade de dados, gerenciamento de metadados, integração de dados, etc.).

Os Dados de Referência também têm características que os distinguem de outros tipos de Dados Mestres (por exemplo, dados de estrutura empresarial e transacional). Eles são menos voláteis. Os conjuntos de Dados de Referência são geralmente menos complexos e menores do que os conjuntos de Dados Mestres ou Transacionais. Eles têm menos colunas e menos linhas. Os desafios da resolução de entidades não fazem parte do Reference Data Management.

O foco do gerenciamento de dados difere entre Dados de Referência e Dados Mestres:

- **O Master Data Management (MDM)** envolve controle sobre valores e identificadores de Master Data que permitem o uso consistente, em todos os sistemas, dos dados mais precisos e oportunos sobre entidades empresariais essenciais. Os objetivos do MDM incluem garantir a disponibilidade de valores precisos e atuais, ao mesmo tempo em que reduz os riscos associados a identificadores ambíguos (aqueles identificados com mais de uma instância de uma entidade e aqueles que se referem a mais de uma entidade).
- **Reference Data Management (RDM)** envolve controle sobre valores de domínio definidos e suas definições. O objetivo do RDM é garantir que a organização tenha acesso a um conjunto completo de valores precisos e atuais para cada conceito representado.

Um desafio do Reference Data Management é o da propriedade ou responsabilidade pela definição e manutenção. Alguns Reference Data se originam fora das organizações que os usam. Alguns cruzam limites organizacionais internos e podem não ser propriedade de um único departamento. Outros Reference Data podem ser criados e mantidos dentro de um departamento, mas têm valor potencial em outro lugar em um

organização. Determinar a responsabilidade pela obtenção de dados e gerenciar atualizações faz parte do RDM. A falta de responsabilização introduz risco, pois diferenças em Dados de Referência podem causar mal-entendidos do contexto dos dados (como quando duas unidades de negócios têm valores diferentes para classificar o mesmo conceito).

Como os Dados Mestre e de Referência fornecem contexto para transações, eles moldam os dados de Transação que entram em uma organização durante as operações (por exemplo, em sistemas CRM e ERP). Eles também enquadram a análise realizada em Dados de Transação.

1.3.2 Dados de Referência

Conforme observado, *Dados de Referência* são quaisquer dados usados para caracterizar ou classificar outros dados, ou para relacionar dados a informações externas a uma organização (Chisholm, 2001). Os Dados de Referência mais básicos consistem em códigos e descrições, mas alguns Dados de Referência 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). Informações externas podem incluir informações geográficas ou padrões (por exemplo, Código do País: DE, US, TR).

Dados de referência podem ser armazenados de diferentes maneiras para atender às diferentes necessidades. Por exemplo, integração de dados (por exemplo, mapeamentos de dados para padronização ou verificações de qualidade de dados) ou outra funcionalidade de aplicativo (por exemplo, anéis de sinônimos para habilitar pesquisa e descoberta). Também pode ter considerações de interface de usuário específicas do dispositivo (por exemplo, vários idiomas). Técnicas comuns de armazenamento usam:

- Tabelas de código em bancos de dados relacionais, vinculadas por meio de chaves estrangeiras a outras tabelas para manter funções de integridade referencial dentro do sistema de gerenciamento de banco de dados
- Sistemas de gerenciamento de dados de referência que mantêm entidades comerciais, valores permitidos, de estado futuro ou obsoletos e regras de mapeamento de termos para oferecer suporte a um uso mais amplo de aplicativos e integração de dados

- Metadados específicos de atributos de objeto para especificar valores permitidos com foco no acesso à API ou à interface do usuário

O *Reference Data Management* envolve o controle e a manutenção de valores de domínio definidos, definições e relacionamentos dentro e entre valores de domínio. O objetivo do Reference Data Management é garantir que os valores sejam consistentes e atuais em diferentes funções e que os dados sejam acessíveis à organização. Como outros dados, os Reference Data exigem Metadados. Um atributo importante de Metadados para Reference Data inclui sua fonte. Por exemplo, o órgão regulador para Reference Data padrão da indústria.

1.3.2.1 Estrutura de Dados de Referência

Dependendo da granularidade e complexidade do que os Dados de Referência representam, eles podem ser estruturados como uma lista simples, uma referência cruzada ou uma taxonomia. A capacidade de usar e manter Dados de Referência deve ser considerada ao estruturá-los em um banco de dados ou em um sistema de Gerenciamento de Dados de Referência.

1.3.2.1.1 Listas

A forma mais simples de Dados de Referência combina um valor de código com uma descrição em uma lista, como na Tabela 17. O valor de código é o identificador primário, o valor de referência de forma abreviada que aparece em outros contextos. A descrição declara o que o código representa. A descrição pode ser exibida no lugar do código em telas, páginas, listas suspensas e relatórios. Observe que, neste exemplo, o valor de código para Reino Unido é GB de acordo com os padrões internacionais, e não UK, embora UK seja uma forma abreviada comum usada em muitas formas de comunicação. Equilíbrio entre conformidade com os padrões e usabilidade ao definir os requisitos de Dados de Referência.

Tabela 17 Lista de referência simples

Table 17 Simple Reference List

Code Value	Description
US	United States of America
GB	United Kingdom (Great Britain)

Dependendo do conteúdo e da complexidade dos Dados de Referência, atributos adicionais podem ser necessários para definir o significado do código. As definições fornecem informações que o rótulo sozinho não fornece. As definições raramente aparecem em relatórios ou listas suspensas.

No entanto, eles aparecem em lugares como funções de Ajuda para aplicativos, que orientam o uso apropriado de códigos em contexto.

Listas, como quaisquer Dados de Referência, devem atender aos requisitos dos consumidores de dados, incluindo requisitos para o nível apropriado de detalhes. Se uma lista de valores for destinada a dar suporte à classificação de dados por usuários casuais, uma lista altamente detalhada provavelmente causará problemas de qualidade de dados e desafios de adoção. Da mesma forma, uma lista de valores muito genérica impediria que os trabalhadores do conhecimento capturassem nível suficiente de detalhes. Para acomodar esses casos, é melhor manter listas distintas que sejam relacionadas em vez de tentar ter uma única lista que seja o padrão para todas as comunidades de usuários. A Tabela 18 fornece um exemplo relacionado a códigos de status para tickets de help desk. Sem as informações fornecidas pela definição, o status do ticket seria ambíguo para qualquer pessoa não familiarizada com o sistema. Essa diferenciação é especialmente necessária para classificações que direcionam métricas de desempenho ou outras análises de Business Intelligence.

Tabela 18 Lista de referência simples Expandido

Table 18 Simple Reference List Expanded

Code	Description	Definition
1	New	Indicates a newly created ticket without an assigned resource
2	Assigned	Indicates a ticket that has a named resource assigned
3	Work In Progress	Indicates the assigned resource started working on the ticket
4	Resolved	Indicates request is assumed to be fulfilled per the assigned resource
5	Cancelled	Indicates request was cancelled based on requester interaction
6	Pending	Indicates request cannot proceed without additional information
7	Fulfilled	Indicates request was fulfilled and verified by the requester

1.3.2.1.2 Listas de Referência Cruzada

Diferentes aplicativos podem usar diferentes conjuntos de códigos para representar o mesmo conceito. Esses conjuntos de códigos podem estar em granularidades diferentes ou na mesma granularidade com valores diferentes. Conjuntos de dados de referência cruzada traduzem entre valores de códigos. A Tabela 19 apresenta uma referência cruzada do Código de Estado dos EUA (um exemplo de múltiplas representações no mesmo nível de granularidade). Os Códigos de Estado do Serviço Postal dos EUA são códigos alfabéticos de dois caracteres. O FIPS usa um numérico para expressar o mesmo conceito. O Código de Estado ISO também inclui uma referência ao país.

Tabela 19 Lista de referências cruzadas

Table 19 Cross-Reference List

USPS State Code	ISO State Code	FIPS Numeric State Code	State Abbreviation	State Name	Formal State Name
CA	US-CA	06	Calif.	California	State of California
KY	US-KY	21	Ky.	Kentucky	Commonwealth of Kentucky
WI	US-WI	55	Wis.	Wisconsin	State of Wisconsin

Os requisitos de idioma podem afetar a estrutura dos Dados de Referência. Listas multilíngues são uma instância específica de uma lista de referência cruzada. Enquanto

listas de códigos fornecem um formato padrão legível por máquina, glossários específicos de idioma fornecem conteúdo utilizável. A Tabela 20 fornece um exemplo do padrão ISO 3166. Existem diferentes maneiras de lidar com listas multilíngues, dependendo de quantos idiomas e conjuntos de caracteres estão envolvidos. As listas não precisam ser normalizadas para serem eficazes. A estrutura desnormalizada torna um pouco mais fácil compreender os relacionamentos.

Tabela 20 Multi-idioma

Lista de Referências

Table 20 Multi-Language Reference List

ISO 3166-1 Alpha 2 Country Code	English Name	Local Name	Local Name Local Alphabet	French Name	...
CN	China	Zhong Guo	中国/中國	Chine	

1.3.2.1.3 Taxonomias

Estruturas de dados de referência taxonômica capturam informações em diferentes níveis de especificidade. Por exemplo, um código postal dos EUA pode ser uma categoria significativa por si só e existe dentro de uma cidade, um condado e um estado. Esses relacionamentos podem ser expressos dentro de uma tabela de referência e vários níveis de análise podem ser feitos usando o código postal como um driver.

As taxonomias permitem a classificação de conteúdo e a navegação multifacetada para dar suporte à Business Intelligence. Os Dados de Referência Taxonômicos podem ser armazenados em um relacionamento recursivo. As ferramentas de gerenciamento de taxonomia também mantêm informações hierárquicas. A Tabela 21 e a Tabela 22 mostram exemplos de duas taxonomias hierárquicas comuns. Em ambos os casos, a hierarquia inclui um código, uma descrição e uma referência a um código pai que classifica os códigos individuais.

Por exemplo, na Tabela 21, Plantas florais (10161600) é um código pai para Rosas, Poinsetias e Orquídeas. Na Tabela 22, Comércio varejista (440000) é o pai para Lojas de alimentos e bebidas (445000), que é o pai

para lojas de alimentos especializados (445200).

Tabela 21 UNSPSC (Classificação Universal de Produtos e Serviços Padrão)⁵⁷

--

Table 21 UNSPSC (Universal Standard Products and Services Classification)⁵⁷

Code Value	Description	Parent Code
10161600	Floral plants	10160000
10161601	Rose plants	10161600
10161602	Poinsettias plants	10161600
10161603	Orchid plants	10161600
10161700	Cut flowers	10160000
10161705	Cut roses	10161700

Tabela 22 NAICS (Sistema de Classificação Industrial da América do Norte)⁵⁸

Table 22 NAICS (North America Industry Classification System)⁵⁸

Code Value	Description	Parent Code
440000	Retail Trade	440000
445000	Food and Beverage Stores	440000
445200	Specialty Food Stores	445000
445210	Meat Markets	445200
445220	Fish and Seafood Markets	445200
445290	Other Specialty Food Stores	445200
445291	Baked Goods Stores	445290
445292	Confectionary and Nut Stores	445290

1.3.2.1.4 Ontologias Algumas

organizações incluem ontologias usadas para gerenciar conteúdo de site como parte dos Dados de Referência. Elas se encaixam nessa categoria, pois são usadas para caracterizar outros dados ou para relacionar dados organizacionais a informações além dos limites da organização.

Ontologias também podem ser entendidas como uma forma de Metadados. Ontologias e outras taxonomias complexas precisam ser gerenciadas de maneira semelhante à forma como os Dados de Referência são gerenciados. Os valores precisam ser completos, atuais e

claramente definido. As melhores práticas para manter ontologias são semelhantes às do Reference Data Management. Um dos principais casos de uso para ontologias é o gerenciamento de conteúdo. Eles são descritos em mais detalhes no [Capítulo 9](#).

1.3.2.2 Dados de Referência Internos ou Proprietários Muitas

organizações criam Dados de Referência para dar suporte a processos e aplicativos internos. Frequentemente, esses dados de referência proprietários geralmente crescem organicamente ao longo do tempo. Parte do RDM inclui gerenciar esses conjuntos de dados e, idealmente, criar consistência entre eles, onde essa consistência serve à organização. Por exemplo, se diferentes unidades de negócios usam termos diferentes para descrever o status de uma conta, então é difícil para qualquer pessoa na organização determinar o número geral de clientes que ela atende em um ponto no tempo. Ao ajudar a gerenciar conjuntos de Dados de Referência internos, os Administradores de Dados devem equilibrar entre a necessidade de ter palavras comuns para as mesmas informações e a necessidade de flexibilidade quando os processos diferem entre si.

1.3.2.3 Dados de Referência da Indústria

Dados de Referência da Indústria é um termo amplo para descrever conjuntos de dados que são criados e mantidos por associações da indústria ou órgãos governamentais, em vez de organizações individuais, a fim de fornecer um padrão comum para codificar conceitos importantes. Essa codificação leva a uma maneira comum de entender dados e é um pré-requisito para compartilhamento e interoperabilidade de dados. Por exemplo, os códigos da Classificação Internacional de Doenças (CID) fornecem uma maneira comum de classificar condições de saúde (diagnósticos) e tratamentos (procedimentos) e, portanto, ter uma abordagem consistente para fornecer cuidados de saúde e entender resultados. Se cada médico e hospital criasse seu próprio conjunto de códigos para doenças, seria virtualmente impossível entender tendências e padrões.

Os Dados de Referência do Setor são produzidos e mantidos externamente às organizações que os usam, mas são necessários para entender as transações dentro dessas organizações. Podem ser necessários para dar suporte a esforços específicos de Gerenciamento de Qualidade de Dados (por exemplo, diretórios comerciais de terceiros), cálculos comerciais (por exemplo, taxas de câmbio) ou dados comerciais

aumento (por exemplo, dados de marketing). Esses conjuntos de dados variam amplamente, dependendo da indústria e do conjunto de códigos individuais. ([Consulte o Capítulo 10.](#))

1.3.2.4 Dados geográficos ou geoestatísticos A referência

geográfica ou geoestatística permite a classificação ou análise com base na geografia. Por exemplo, os relatórios do bureau de censo descrevem a densidade populacional e as mudanças demográficas que dão suporte ao planejamento e à pesquisa de mercado. O histórico climático mapeado para uma classificação geográfica rigorosa pode dar suporte ao gerenciamento de inventário e ao planejamento promocional.

1.3.2.5 Dados de Referência Computacionais Muitas

atividades comerciais dependem do acesso a cálculos comuns e consistentes. Por exemplo, cálculos de câmbio estrangeiro dependem de tabelas de valores de câmbio gerenciadas e com registro de data e hora. Os Dados de Referência Computacionais diferem de outros tipos devido à frequência com que mudam. Muitas organizações compram esse tipo de dados de terceiros que garantem que eles sejam completos e precisos.

Tentar manter esses dados internamente provavelmente causará problemas de latência.

1.3.2.6 Metadados do conjunto de dados de referência padrão

Dados de Referência, como outros dados, podem mudar ao longo do tempo. Dada sua prevalência em qualquer organização, é importante manter Metadados-chave sobre conjuntos de Dados de Referência para garantir que sua linhagem e atualidade sejam compreendidas e mantidas. A Tabela 23 fornece exemplos desses Metadados.

Tabela 23 Referência Crítica

Atributos de metadados de dados

Table 23 Critical Reference Data Metadata Attributes

Reference Data Set Key Information	Description
Formal Name	Official, especially if external name of the Reference Data set (e.g., ISO 3166-1991 Country Code List)
Internal Name	Name associated with the data set within the organization (e.g., Country Codes – ISO)
Data Provider	The party that provides and maintains the Reference Data set. This can be external (ISO), internal (a specific department), or external – extended (obtained from an external party but then extended and modified internally).
Data Provider Data Set Source	Description of where data provider's data sets can be obtained. This is likely a Universal Resource Identifier (URI) within or outside of the enterprise network.
Data Provider Latest Version Number	If available and maintained, this describes the latest version of the external data provider's data set where information may be added or deprecated from the version in the organization
Data Provider Latest Version Date	If available and maintained, this describes when the standard list was last updated
Internal Version Number	Version number of the current Reference Data set or version number of the last update that was applied against the data set
Internal Version Reconciliation Date	Date when data set was last updated based on the external source
Internal Version Last Update Date	Date data set was last changed. This does not mean reconciliation with an external version.

1.3.3 Dados Mestres

Master Data são dados sobre as entidades comerciais (por exemplo, funcionários, clientes, produtos, estruturas financeiras, ativos e locais) que fornecem contexto para transações comerciais e análises. Uma entidade é um objeto do mundo real (pessoa, organização, lugar ou coisa). As entidades são representadas por instâncias de entidade, na forma de dados/registros. Master

Os dados devem representar os dados mais confiáveis e precisos disponíveis sobre as principais entidades comerciais. Quando bem gerenciados, os valores do Master Data são confiáveis e podem ser usados com segurança.

As regras de negócios normalmente ditam o formato e os intervalos permitidos de valores de Master Data. Master Data organizacional comum inclui dados sobre:

- **Partes**, compostas por indivíduos e organizações, e suas funções, como clientes, cidadãos, pacientes, fornecedores, agentes, parceiros de negócios, concorrentes, funcionários ou alunos
- **Produtos e serviços**, tanto internos quanto externos
- **Estruturas financeiras**, como contratos, contas contábeis, centros de custo ou centros de lucro
- **Locais**, como endereços e coordenadas GPS

1.3.3.1 Sistema de Registro, Sistema de Referência Quando há

versões potencialmente diferentes da "verdade", é necessário distingui-las. Para fazer isso, é preciso saber onde os dados se originam ou são acessados, e quais dados foram preparados para usos específicos. Um *Sistema de Registro* é um sistema autoritativo onde os dados são criados/capturados e/ou mantidos por meio de um conjunto definido de regras e expectativas (por exemplo, um sistema ERP pode ser o Sistema de Registro para clientes de venda). Um *Sistema de Referência* é um sistema autoritativo onde os consumidores de dados podem obter dados confiáveis para dar suporte a transações e análises, mesmo que as informações não tenham se originado no sistema de referência. Aplicativos MDM, Hubs de Compartilhamento de Dados e Data Warehouses geralmente servem como sistemas de referência.

1.3.3.2 Fonte Confiável, Registro Dourado Uma

Fonte Confiável é reconhecida como a 'melhor versão da verdade' com base em uma combinação de regras automatizadas e administração manual do conteúdo de dados. Uma fonte confiável também pode ser chamada de Visão Única,

Visão 360°. Qualquer sistema MDM deve ser gerenciado para que seja uma fonte confiável. Dentro de uma fonte confiável, os registros que representam os dados mais precisos sobre instâncias de entidade podem ser chamados de *Golden Records*.

O termo *Golden Record* pode ser enganoso. A Tech Target define um Golden Record como "a 'versão única da verdade', onde 'verdade' é entendida como a referência à qual os usuários de dados podem recorrer quando desejam garantir que têm a versão correta de uma informação. O golden record abrange todos os dados em cada sistema de registro (SOR) dentro de uma organização específica."59 No entanto, as duas partes desta definição colocam o conceito em questão, pois os dados em diferentes sistemas podem não se alinhar em 'uma única versão da verdade'.

Em qualquer esforço de Dados Mestres, a fusão/resolução de dados de várias fontes em um "Registro Dourado" não significa que ele seja sempre uma representação 100% completa e 100% precisa de todas as entidades dentro da organização (especialmente em organizações que têm vários SORs fornecendo dados para o ambiente de Dados Mestres).

Prometer que os dados são "de ouro" quando não são pode minar a confiança dos consumidores de dados.

É por isso que alguns preferem o termo *Trusted Source* para se referir à "melhor versão que temos" dos Master Data. Fazer isso coloca ênfase em como os dados são definidos e gerenciados para chegar a uma melhor versão. Também ajuda diferentes consumidores de dados a verem as peças componentes da "versão única" que são importantes para eles. As áreas de Finanças e Atuariais geralmente têm uma perspectiva diferente da "versão única" do Cliente do que a área de Marketing. A *Trusted Source* fornece múltiplas perspectivas de entidades comerciais conforme identificadas e definidas pelos Data Stewards.

1.3.3.3 Gerenciamento de Dados Mestres Conforme

descrito na introdução do capítulo, o Gerenciamento de Dados Mestres envolve o controle sobre os valores e identificadores de Dados Mestres que permitem o uso consistente, em todos os sistemas, dos dados mais precisos e oportunos sobre entidades empresariais essenciais. Os objetivos incluem garantir

disponibilidade de valores precisos e atuais, reduzindo ao mesmo tempo o risco de identificadores ambíguos.

A Gartner define o Master Data Management como “uma disciplina habilitada pela tecnologia na qual os negócios e a TI trabalham juntos para garantir a uniformidade, precisão, administração, consistência semântica e responsabilidade dos ativos oficiais de Master Data compartilhados da empresa.

Dados Mestres são o conjunto consistente e uniforme de identificadores e atributos estendidos que descrevem as principais entidades da empresa, incluindo clientes, clientes potenciais, cidadãos, fornecedores, sites, hierarquias e plano de contas.”⁶⁰

A definição da Gartner enfatiza que o MDM é uma disciplina, composta de pessoas, processos e tecnologia. Não é uma solução de aplicativo específica. Infelizmente, a sigla MDM (Master Data Management) é frequentemente usada para se referir a sistemas ou produtos usados para gerenciar Master Data.⁶¹ Os aplicativos MDM podem facilitar os métodos, e às vezes de forma bastante eficaz, mas usar um aplicativo MDM não garante que o Master Data esteja sendo gerenciado para atender às necessidades organizacionais.

Avaliar os requisitos de MDM de uma organização inclui identificar:

- Quais papéis, organizações, lugares e coisas são referenciados repetidamente
- Quais dados são usados para descrever pessoas, organizações, lugares e coisas
- Como os dados são definidos e estruturados, incluindo a granularidade dos dados
- Onde os dados são criados/obtidos, armazenados, disponibilizados e acessados
- Como os dados mudam à medida que se movem pelos sistemas dentro da organização
- Quem usa os dados e para quais propósitos
- Quais critérios são usados para entender a qualidade e confiabilidade dos dados e suas fontes

O Master Data Management é desafiador. Ele ilustra um desafio fundamental com dados: as pessoas escolhem maneiras diferentes de representar conceitos semelhantes e a reconciliação entre essas representações nem sempre é direta; tão importante quanto isso, as informações mudam ao longo do tempo e contabilizar sistematicamente essas mudanças exige planejamento, conhecimento de dados e habilidades técnicas. Em suma, dá trabalho.

Qualquer organização que tenha reconhecido a necessidade de MDM provavelmente já tem um cenário de sistema complexo, com várias maneiras de capturar e armazenar referências a entidades do mundo real. Devido ao crescimento orgânico ao longo do tempo ou a fusões e aquisições, os sistemas que forneceram entrada para o processo de MDM podem ter definições diferentes das próprias entidades e muito provavelmente ter padrões diferentes para Qualidade de Dados. Devido a essa complexidade, é melhor abordar o Master Data Management um domínio de dados por vez. Comece pequeno, com um punhado de atributos, e desenvolva ao longo do tempo.

O planejamento do gerenciamento de dados mestres inclui várias etapas básicas. Dentro de um domínio:

- Identificar fontes candidatas que fornecerão uma visão abrangente das entidades de Dados Mestres
- Desenvolver regras para combinar e mesclar com precisão instâncias de entidades
- Estabelecer uma abordagem para identificar e restaurar dados inadequadamente combinados e mesclados
- Estabelecer uma abordagem para distribuir dados confiáveis para sistemas em toda a empresa

Executar o processo, no entanto, não é tão simples quanto essas etapas sugerem, pois o MDM é um processo de gerenciamento de ciclo de vida. Atividades críticas para o ciclo de vida incluem:

- Estabelecer o contexto das entidades de Dados Mestres, incluindo definições de atributos associados e as condições de

seu uso. Este processo requer governança.

- Identificar múltiplas instâncias da mesma entidade representadas dentro e entre fontes de dados; criar e manter identificadores e referências cruzadas para permitir a integração de informações.
- Reconciliar e consolidar dados entre fontes para fornecer um registro mestre ou a melhor versão da verdade.
Registros consolidados fornecem uma visão unificada das informações entre sistemas e buscam resolver inconsistências na nomenclatura de atributos e no valor dos dados.
- Identificar instâncias incorretamente correspondidas ou mescladas e garantir que elas sejam resolvidas e corretamente associadas aos identificadores.
- Provisionamento de acesso a dados confiáveis em todos os aplicativos, seja por meio de leituras diretas, serviços de dados ou por feeds de replicação para dados transacionais, de armazenamento ou analíticos lojas.
- Reforçar o uso dos valores de Master Data dentro da organização. Este processo também requer governança e gerenciamento de mudanças para assegurar uma perspectiva empresarial compartilhada.

1.3.3.4 Principais etapas de processamento do gerenciamento de dados

mestres As principais etapas de processamento para MDM são ilustradas na Figura 76. Elas incluem gerenciamento de modelo de dados; aquisição de dados; validação, padronização e enriquecimento de dados; resolução de entidades; e administração e compartilhamento.

Em um ambiente MDM abrangente, o modelo de dados lógicos será fisicamente instanciado em múltiplas plataformas. Ele orienta a implementação da solução MDM, fornecendo a base dos serviços de integração de dados. Ele deve orientar como os aplicativos são configurados para aproveitar as vantagens dos recursos de reconciliação de dados e verificação da qualidade dos dados.

Figura 76 Processamento de Chaves

Etapas para MDM

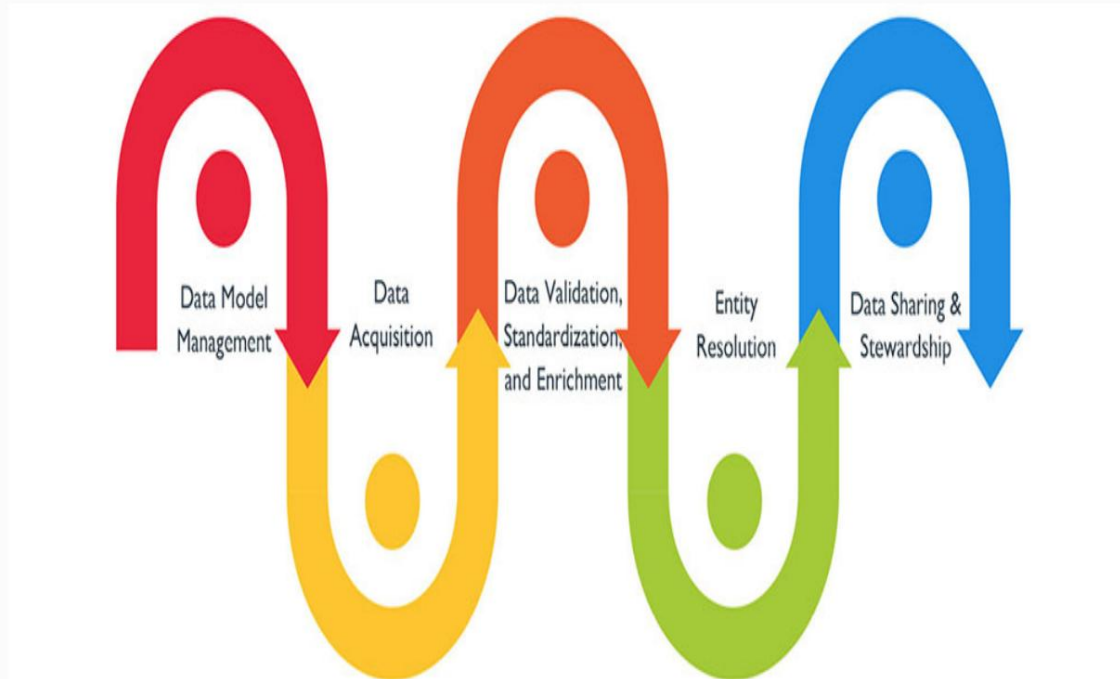


Figure 76 Key Processing Steps for MDM

1.3.3.4.1 Gerenciamento de Modelo de Dados

O trabalho de Dados Mestres traz à tona a importância de definições de dados lógicas claras e consistentes. O modelo deve ajudar a organização a superar a "linguagem do sistema". Termos e definições usados em um sistema de origem podem fazer sentido dentro dos limites desse sistema, mas nem sempre fazem sentido em um nível empresarial. Para Dados Mestres, termos e definições usados em um nível empresarial devem estar no contexto dos negócios conduzidos em toda a organização e não necessariamente dependentes do sistema de origem contribuindo com valores de dados.

Para atributos que compõem o Master Data, a granularidade da definição e os valores de dados associados também devem fazer sentido em toda a organização.

Os sistemas de origem podem apresentar o mesmo nome de atributo, mas os valores de dados estão em contextos completamente diferentes no

nível empresarial. Da mesma forma, os sistemas de origem podem apresentar atributos com nomes diferentes que, no nível empresarial, coalescem em um único atributo e os valores de dados estão no contexto apropriado. Às vezes, vários atributos são apresentados de uma única fonte e seus respectivos valores de dados são usados para derivar um único valor de dados para um atributo definido no nível empresarial.

1.3.3.4.2 Aquisição de Dados Mesmo

dentro de uma determinada fonte, os dados que representam a mesma instância de entidade podem parecer diferentes, conforme ilustrado na Tabela 24, onde há inconsistências em como nomes, endereços e números de telefone são apresentados. Este exemplo será referenciado novamente mais adiante no capítulo.

Tabela 24 Dados de origem como
Recebido pelo Sistema MDM

Table 24 Source Data as Received by the MDM System

Source ID	Name	Address	Telephone
123	John Smith	123 Main, Dataland, SQ 98765	
234	J. Smith	123 Main, Dataland, DA	2345678900
345	Jane Smith	123 Main, Dataland, DA	234-567-8900

O planejamento, avaliação e incorporação de novas fontes de dados na solução Master Data Management deve ser um processo confiável e repetível. As atividades de aquisição de dados envolvem:

- Receber e responder a novas solicitações de aquisição de fontes de dados
- Realizar avaliações rápidas, ad hoc, de correspondência e de alto nível da qualidade dos dados usando ferramentas de limpeza e criação de perfil de dados
- Avaliar e comunicar a complexidade da integração de dados aos solicitantes para ajudá-los com seus custos

análise de benefícios

- Pilotagem da aquisição de dados e seu impacto nas regras da partida
- Finalizando métricas de qualidade de dados para a nova fonte de dados
- Determinar quem será responsável por monitorar e manter a qualidade dos dados de uma nova fonte
- Concluindo a integração no ambiente geral de gerenciamento de dados

1.3.3.4.3 Validação, Padronização e Enriquecimento de Dados Para permitir a resolução de entidades, os dados devem ser tornados o mais consistentes possível. Isso implica, no mínimo, reduzir a variação no formato e reconciliar valores. Dados de entrada consistentes reduzem a chance de erros na associação de registros. Os processos de preparação incluem:

- **Validação:** Identificar dados comprovadamente errôneos ou provavelmente incorretos ou inadimplentes (por exemplo, remoção de endereços de e-mail claramente falsos)
- **Padronização:** Garantir que o conteúdo dos dados esteja em conformidade com os valores de Dados de Referência padrão (por exemplo, códigos de país), formatos (por exemplo, números de telefone) ou campos (por exemplo, endereços)
- **Enriquecimento:** Adicionar atributos que podem melhorar os serviços de resolução de entidades (por exemplo, Dunn and Bradstreet DUNS Number e Ultimate DUNS Number para registros de empresas relacionadas, Acxiom ou Experian Consumer IDs para registros individuais)

A Tabela 25 ilustra os resultados do processo de limpeza e padronização no exemplo da Tabela 24. Endereços que tinham formatos diferentes agora são reconhecidamente os mesmos. Números de telefone incluem formatação padrão.

Tabela 25 Padronizada e

Dados de entrada enriquecidos

Table 25 Standardized and Enriched Input Data

Source ID	Name	Address (Cleansed)	Telephone (Cleansed)
123	John Smith	123 Main, Dataland, SQ 98765	
234	J. Smith	123 Main, Dataland, SQ 98765	+1 234 567 8900
345	Jane Smith	123 Main, Dataland, SQ 98765	+1 234 567 8900

1.3.3.4.4 Resolução de Entidade e Gerenciamento de Identificadores

Resolução de entidade é o processo de determinar se duas referências a objetos do mundo real se referem ao mesmo objeto ou a objetos diferentes (Talburt, 2011).

Resolução de entidade é um processo de tomada de decisão. Modelos para executar o processo diferem com base na abordagem que eles tomam para determinar similaridade entre duas referências.

Embora a resolução sempre ocorra entre pares de referências, o processo pode ser sistematicamente estendido para incluir grandes conjuntos de dados.

A resolução de entidades é essencial para o MDM, pois o processo de correspondência e mesclagem de registros permite a construção do conjunto de dados mestres.

A resolução de entidades inclui um conjunto de atividades (extração de referência, preparação de referência, resolução de referência, gerenciamento de identidade, análise de relacionamento) que permitem que a identidade de instâncias de entidade e o relacionamento entre instâncias de entidade sejam gerenciados ao longo do tempo.

Dentro do processo de resolução de referência, duas referências podem ser identificadas como representando a mesma entidade, por meio do processo de determinação de equivalência. Essas referências podem então ser vinculadas por meio de um valor (um identificador global) que indica que elas são equivalentes (Talburt, 2011).

1.3.3.4.4.1 Correspondência

A *correspondência*, ou identificação de candidatos, é o processo de identificar como diferentes registros podem se relacionar a uma única entidade. Os riscos com isso

processo são:

- **Falsos positivos:** Duas referências que não representam a mesma entidade são vinculadas a um único identificador. Isso resulta em um identificador que se refere a mais de uma instância de entidade do mundo real.
- **Falsos negativos:** Duas referências representam a mesma entidade, mas não estão vinculadas a um único identificador. Isso resulta em vários identificadores que se referem à mesma entidade do mundo real quando se espera que cada instância tenha um e somente um identificador.

Ambas as situações são abordadas por meio de um processo chamado *análise de similaridade* ou *correspondência*, no qual o grau de similaridade entre quaisquer dois registros é pontuado, geralmente com base na correspondência aproximada ponderada entre valores de atributos correspondentes. Se a pontuação estiver acima de um limite especificado, os dois registros são considerados como representantes da mesma entidade (uma correspondência). Por meio da análise de similaridade, pequenas variações nos dados podem ser reconhecidas e os valores dos dados podem ser consolidados. Duas abordagens básicas, que podem ser usadas juntas, são determinística e probabilística:

- Algoritmos **determinísticos**, como análise sintática e padronização, dependem de padrões e regras definidos para atribuir pesos e pontuações para determinar similaridade. Algoritmos determinísticos são previsíveis, pois os padrões correspondidos e as regras aplicadas sempre produzirão os mesmos resultados. Esse tipo de correspondência funciona imediatamente com desempenho relativamente bom, mas é tão bom quanto as situações previstas pelas pessoas que desenvolveram as regras.
- Algoritmos **probabilísticos** dependem de técnicas estatísticas para avaliar a probabilidade de que qualquer par de registros represente a mesma entidade. Isso depende da capacidade de obter dados

amostras para fins de treinamento, observando os resultados esperados para um subconjunto dos registros e ajustando o matcher para se autoajustar com base na análise estatística. Esses matchers não dependem de regras, então os resultados podem ser não determinísticos.

No entanto, como as probabilidades podem ser refinadas com base na experiência, os correspondentes probabilísticos conseguem melhorar sua precisão de correspondência à medida que mais dados são analisados.

1.3.3.4.4.2 Resolução de identidade

Algumas correspondências ocorrem com grande confiança, com base em correspondências exatas de dados em vários campos. Outras correspondências são sugeridas com menos confiança devido a valores conflitantes. Por exemplo:

- Se dois registros compartilham o mesmo sobrenome, nome, data de nascimento e número de previdência social, mas o endereço da rua é diferente, é seguro presumir que eles se referem à mesma pessoa que mudou seu endereço postal?
- Se dois registros compartilham o mesmo número de previdência social, endereço residencial e primeiro nome, mas o sobrenome é diferente, é seguro assumir que eles se referem à mesma pessoa que mudou seu sobrenome? A probabilidade aumentaria ou diminuiria com base no gênero e na idade?
- Como esses exemplos mudam se o número de previdência social for desconhecido para um registro? Quais outros identificadores são úteis para determinar a probabilidade de uma correspondência? Quanta confiança é necessária para a organização afirmar uma correspondência?

A Tabela 26 ilustra a conclusão do processo para os registros de amostra na Tabela 24 e Tabela 25. Aqui, as duas segundas instâncias de entidade (ID de origem 234 e 345) são determinadas como representando a mesma pessoa (Jane Smith), enquanto a primeira (ID de origem 123) é identificada como representando uma pessoa diferente (John Smith).

Tabela 26 Candidato Identificação e Identidade Resolução

Table 26 Candidate Identification and Identity Resolution

Source ID	Name	Address (Cleansed)	Telephone (Cleansed)	Candidate ID	Party ID
123	John Smith	123 Main, Dataland, SQ 98765		XYZ	1
234	J. Smith	123 Main, Dataland, SQ 98765	+1 234 567 8900	XYZ, ABC	2
345	Jane Smith	123 Main, Dataland, SQ 98765	+1 234 567 8900	ABC	2

Apesar dos melhores esforços, as decisões de correspondência às vezes se mostram incorretas. É essencial manter o histórico de correspondências para que as correspondências possam ser desfeitas quando forem descobertas incorretas. As métricas de taxa de correspondência permitem que as organizações monitorem o impacto e a eficácia de suas regras de inferência de correspondência. O reprocessamento de regras de correspondência pode ajudar a identificar melhores candidatos de correspondência à medida que novas informações são recebidas pelo processo de resolução da entidade.

1.3.3.4.4.3 Correspondência de fluxos de trabalho/tipos de reconciliação

Regras de correspondência para diferentes cenários exigem fluxos de trabalho diferentes:

- **As regras de correspondência de identificação duplicada** focam em um conjunto específico de elementos de dados que identificam exclusivamente uma entidade e identificam oportunidades de mesclagem sem tomar ação automática. Os Business Data Stewards podem revisar essas ocorrências e decidir tomar ação caso a caso.
- **As regras de link de correspondência** identificam e fazem referência cruzada a registros que parecem se relacionar a um registro mestre sem atualizar o conteúdo do registro com referência cruzada. As regras de link de correspondência são

mais fácil de implementar e muito mais fácil de reverter.

- **Regras de correspondência-mesclagem** correspondem a registros e mesclam os dados desses registros em um único registro unificado, reconciliado e abrangente. Se as regras se aplicarem a todas as fontes de dados, crie um único registro exclusivo e abrangente em cada repositório de dados. No mínimo, use dados confiáveis de um repositório de dados para complementar dados em outros repositórios de dados, substituindo valores ausentes ou valores considerados imprecisos.

As regras de correspondência e mesclagem são complexas e buscam fornecer a versão unificada e reconciliada das informações em vários registros e fontes de dados. A complexidade se deve à necessidade de identificar qual campo de qual fonte pode ser confiável com base em uma série de regras. A introdução de cada nova fonte pode alterar essas regras ao longo do tempo.

Os desafios com regras de correspondência e mesclagem incluem a complexidade operacional de reconciliar os dados e o custo de reverter a operação se houver uma mesclagem falsa.

A vinculação de correspondência é uma operação mais simples, pois atua no registro de referência cruzada e não nos atributos individuais do registro de Dados Mestres mesclado, embora possa ser mais difícil apresentar informações abrangentes de vários registros.

Reavalie periodicamente as regras de match-merge e match-link porque os níveis de confiança mudam ao longo do tempo. Muitos mecanismos de correspondência de dados fornecem correlações estatísticas de valores de dados para ajudar a estabelecer níveis de confiança. (Consulte [o Capítulo 13.](#))

1.3.3.4.4 Gerenciamento de ID de Dados Mestres

O gerenciamento de Dados Mestres envolve o gerenciamento de identificadores. Há dois tipos de identificadores que precisam ser gerenciados em todas as fontes de dados em um ambiente MDM: IDs globais e informações de referência cruzada (x-Ref).

Um *ID global* é o identificador exclusivo atribuído e mantido pela solução MDM anexado aos registros reconciliados. Sua finalidade é identificar exclusivamente a instância da entidade. No exemplo da Tabela 26, quando

vários registros foram determinados como representando a mesma instância de entidade; o valor 'ABC' foi atribuído a ambos como um ID candidato.

Os registros foram resolvidos para o ID de parte única '2'.

IDs globais devem ser gerados por apenas uma solução autorizada, independentemente de qual tecnologia esteja executando atividades de integração de Master Data, para evitar qualquer risco de valores duplicados. IDs globais podem ser números ou GUIDs (Global Unique Identifiers), desde que a exclusividade possa ser mantida. A principal complexidade que precisa ser tratada para a geração de ID global é como manter o ID global correto (para executar atualizações de dados downstream apropriadas) devido a um unmerge-remerge. *O gerenciamento de X-Ref* é o gerenciamento do relacionamento entre IDs de origem e o ID global. O gerenciamento de X-Ref deve incluir recursos para manter o histórico de tais mapeamentos para dar suporte a métricas de taxa de correspondência e expor serviços de pesquisa para permitir a integração de dados.

1.3.3.4.4.5 Gerenciamento de Afiliação

Gerenciamento de Afiliação é estabelecer e manter relacionamentos entre registros de Dados Mestres de entidades que têm relacionamentos no mundo real. Exemplos incluem afiliações de propriedade (por exemplo, a Empresa X é uma subsidiária da Empresa Y, um relacionamento pai-filho) ou outras associações (por exemplo, a Pessoa XYZ trabalha na Empresa X).

O design da arquitetura de dados de uma solução MDM deve decidir se deve aproveitar relacionamentos pai-filho, relacionamentos de afiliação ou ambos para uma determinada entidade.

- **Os relacionamentos de afiliação** fornecem a maior flexibilidade por meio da lógica de programação. O tipo de relacionamento pode ser usado para expor esses dados em uma hierarquia pai-filho. Muitas soluções downstream, como ferramentas de relatórios ou navegação de conta, gostariam de ver uma visão hierárquica das informações.
- **Os relacionamentos pai-filho** exigem menos lógica de programação, pois a estrutura de navegação é implícita. No entanto, se o

mudanças no relacionamento e não há uma estrutura de afiliação disponível, isso pode influenciar a qualidade dos dados e as dimensões de Business Intelligence.

1.3.3.4.5 Compartilhamento e administração de dados

Embora grande parte do trabalho do Master Data Management possa ser automatizado por meio de ferramentas que permitem o processamento de grandes números de registros, ele ainda requer administração para resolver situações em que os dados são correspondidos incorretamente. Idealmente, as lições aprendidas do processo de administração podem ser usadas para melhorar algoritmos de correspondência e reduzir instâncias de trabalho manual. (Consulte os [Capítulos 3](#) e [8](#).)

1.3.3.5 Dados Mestres de Partes

Os *Dados Mestres de Partes* incluem dados sobre indivíduos, organizações e as funções que desempenham em relacionamentos comerciais. No ambiente comercial, as partes incluem clientes, funcionários, fornecedores, parceiros e concorrentes. No setor público, as partes geralmente são cidadãos. A aplicação da lei se concentra em suspeitos, testemunhas e vítimas. As organizações sem fins lucrativos se concentram em membros e doadores.

Enquanto na área da saúde o foco está nos pacientes e provedores, na educação o foco está nos alunos e professores.

Os sistemas de Customer Relationship Management (CRM) gerenciam Master Data sobre clientes. O objetivo do CRM é fornecer informações completas e precisas sobre cada cliente.

Um aspecto essencial do CRM é identificar dados duplicados, redundantes ou conflitantes de diferentes sistemas e determinar se os dados representam um ou mais de um cliente. O CRM deve ser capaz de resolver valores conflitantes, reconciliar diferenças e representar com precisão o conhecimento atual do cliente. Esse processo requer regras robustas, bem como conhecimento da estrutura, granularidade, linhagem e qualidade das fontes de dados.

Os sistemas MDM especializados desempenham funções semelhantes para indivíduos, organizações e suas funções, funcionários e fornecedores. Independentemente do setor ou foco, o gerenciamento de Master Data de terceiros representa uma tarefa única.

desafios:

- A complexidade dos papéis e relacionamentos desempenhados por indivíduos e organizações
- Dificuldades na identificação única
- O número de fontes de dados e as diferenças entre elas
- Os múltiplos canais de comunicação móvel e social
- A importância dos dados
- As expectativas de como os clientes desejam se envolver

O Master Data é particularmente desafiador para partes que desempenham múltiplas funções em uma organização (por exemplo, um funcionário que também é cliente) e que utilizam diferentes pontos de contato ou métodos de engajamento (por exemplo, interação por meio de aplicativo de dispositivo móvel vinculado a um site de mídia social).

1.3.3.6 Dados Mestres Financeiros

O *Financial Master Data* inclui dados sobre unidades de negócios, centros de custo, centros de lucro, contas do razão geral, orçamentos, projeções e projetos. Normalmente, um sistema Enterprise Resource Planning (ERP) serve como o hub central para o Financial Master Data (plano de contas), com detalhes do projeto e transações criadas e mantidas em um ou mais aplicativos spoke. Isso é especialmente comum em organizações com funções de back-office distribuídas.

As soluções de Dados Mestres Financeiros não apenas criam, mantêm e compartilham informações; muitas também podem simular como as alterações nos dados financeiros existentes podem afetar os resultados financeiros da organização. As simulações de Dados Mestres Financeiros geralmente fazem parte dos módulos de relatórios, análises e planejamento de Business Intelligence, bem como orçamentos e projeções mais diretos. Por meio desses aplicativos, versões de estruturas financeiras podem ser modeladas para entender os potenciais impactos financeiros. Uma vez que uma decisão é tomada, o acordado

mudanças estruturais podem ser disseminadas para todos os sistemas apropriados.

1.3.3.7 Dados Mestres Legais *Dados*

Mestres Legais incluem dados sobre contratos, regulamentações e outros assuntos legais.

Dados Mestres Legais permitem a análise de contratos para diferentes entidades que fornecem os mesmos produtos ou serviços, para permitir melhor negociação ou combinar contratos em Contratos Mestres.

1.3.3.8 Dados Mestres do Produto

Product Master Data pode focar em produtos e serviços internos de uma organização ou em produtos e serviços de toda a indústria (incluindo concorrentes). Diferentes tipos de soluções de Product Master Data dão suporte a diferentes funções de negócios.

- **O Product Lifecycle Management (PLM)** foca no gerenciamento do ciclo de vida de um produto ou serviço desde a concepção, passando pelo desenvolvimento, fabricação, venda/entrega, serviço e descarte. As organizações implementam sistemas PLM para reduzir o tempo de lançamento no mercado. Em indústrias com longos ciclos de desenvolvimento de produtos (de 8 a 12 anos na indústria farmacêutica), os sistemas PLM permitem que as organizações rastreiem os custos entre processos e acordos legais à medida que os conceitos de produtos evoluem de ideias para produtos potenciais sob nomes diferentes e acordos de licenciamento potencialmente diferentes.
- **O Product Data Management (PDM)** oferece suporte a funções de engenharia e manufatura capturando e permitindo o compartilhamento seguro de informações do produto, como documentos de design (por exemplo, desenhos CAD), receitas (instruções de manufatura), procedimentos operacionais padrão e listas de materiais. A funcionalidade do PDM pode ser habilitada por meio de sistemas especializados ou aplicativos ERP.
- **Os dados do produto nos sistemas de planejamento de recursos empresariais (ERP)** concentram-se em SKUs para dar suporte à entrada de pedidos até

nível de inventário, onde unidades individuais podem ser identificadas por meio de uma variedade de técnicas.

- **Dados de produtos em Manufacturing Execution Systems (MES)** focam em estoque bruto, produtos semiacabados e produtos acabados, onde produtos acabados se vinculam a produtos que podem ser armazenados e pedidos por meio do sistema ERP. Esses dados também são importantes em toda a cadeia de suprimentos e sistemas de logística.
- **Os dados do produto em um sistema de gerenciamento de relacionamento com o cliente (CRM)** que oferece suporte a interações de marketing, vendas e suporte podem incluir famílias e marcas de produtos, associação de representantes de vendas e gerenciamento de território do cliente, bem como campanhas de marketing.

Muitos mestres de produtos estão intimamente ligados aos sistemas de gerenciamento de dados de referência.

1.3.3.9 Dados Mestres de Localização

O Location Master Data fornece a capacidade de rastrear e compartilhar informações geográficas e criar relacionamentos hierárquicos ou territórios com base em informações geográficas. A distinção entre referência e Master Data fica confusa para dados de localização. Aqui está a diferença:

- **Os Dados de Referência de Localização** geralmente incluem dados geopolíticos, como países, estados ou províncias, condados, cidades ou vilas, códigos postais e coordenadas de posicionamento geográfico, como latitude, longitude e altitude. Esses dados raramente mudam, e as mudanças são tratadas por organizações externas. Os Dados de Referência de Localização também podem incluir regiões geográficas e territórios de vendas, conforme definido pela organização.
- **Os dados mestres de localização** incluem endereços de partes comerciais e locais de partes comerciais, bem como endereços de instalações para

locais de propriedade da organização. Conforme as organizações crescem ou se contraem, esses endereços mudam com mais frequência do que outros Dados de Referência de Localização.

Diferentes indústrias exigem dados especializados de ciências da terra (dados geográficos sobre falhas sísmicas, planícies de inundação, solo, precipitação anual e áreas de risco de clima severo) e dados sociológicos relacionados (população, etnia, renda e risco de terrorismo), geralmente fornecidos por fontes externas. fontes.

1.3.3.10 Dados Mestres do Setor – Diretórios de Referência
Diretórios de Referência são listagens autoritativas de entidades de Dados Mestres (empresas, pessoas, produtos, etc.) que as organizações podem comprar e usar como base para suas transações. Enquanto diretórios de referência são criados por organizações externas, uma versão gerenciada e reconciliada das informações é mantida nos próprios sistemas da organização.

Exemplos de diretórios de referência licenciados incluem o Diretório de Empresas da Dun and Bradstreet (D&B) com sedes, subsidiárias e filiais em todo o mundo, e o Banco de Dados de Prescritores da Associação Médica Americana.

Os diretórios de referência permitem o uso de Dados Mestres por:

- Fornecendo um ponto de partida para combinar e vincular novos registros. Por exemplo, em um ambiente com cinco fontes de dados, cada fonte pode ser comparada com o diretório (5 pontos de comparação) vs. entre si (10 pontos de comparação).
- Fornecer elementos de dados adicionais que podem não estar tão facilmente disponíveis no momento da criação do registro (por exemplo, para um médico, isso pode incluir o status da licença médica; para uma empresa, isso pode incluir uma classificação do setor NAICS de seis dígitos).

À medida que os registros de uma organização correspondem e se reconciliam com os diretórios de referência, o registro confiável se desviará do diretório de referência com rastreabilidade para outros registros de origem, atributos contribuintes e regras de transformação.

1.3.4 Arquitetura de Compartilhamento

de Dados Existem várias abordagens arquitetônicas básicas para referência e integração de Dados Mestres. Cada área de assunto de Dados Mestres provavelmente terá seu próprio sistema de registro. Por exemplo, o sistema de recursos humanos geralmente serve como o sistema de registro para dados de funcionários. Um sistema CRM pode servir como o sistema de registro para dados de clientes, enquanto um sistema ERP pode servir como o sistema de registro para dados financeiros e de produtos.

O modelo de arquitetura de hub de compartilhamento de dados mostrado na Figura 77 representa uma arquitetura hub-and-spoke para Master Data. O hub Master Data pode lidar com interações com itens spoke, como sistemas de origem, aplicativos de negócios e armazenamentos de dados, minimizando o número de pontos de integração. Um hub de dados local pode estender e dimensionar o hub Master Data. (Consulte [o Capítulo 8](#).)

Cada uma das três abordagens básicas para implementar um ambiente de hub de dados mestres tem prós e contras:

- Um **Registro** é um índice que aponta para Dados Mestres nos vários sistemas de registro. Os sistemas de registro gerenciam Dados Mestres locais para seus aplicativos. O acesso aos Dados Mestres vem do índice mestre. Um registro é relativamente fácil de implementar porque requer poucas alterações nos sistemas de registro. Mas, frequentemente, consultas complexas são necessárias para montar Dados Mestres de vários sistemas. Além disso, várias regras de negócios precisam ser implementadas para abordar diferenças semânticas entre sistemas em vários lugares.
- Em um **Hub de Transações**, os aplicativos fazem interface com o hub para acessar e atualizar os Dados Mestres. Os Dados Mestres existem dentro do Hub de Transações e não dentro de nenhum outro

aplicações. O Transaction Hub é o sistema de registro para Master Data. Os Transaction Hubs permitem melhor governança e fornecem uma fonte consistente de Master Data. No entanto, é custoso remover a funcionalidade para atualizar Master Data de sistemas de registro existentes. As regras de negócios são implementadas em um único sistema: o Hub.

- **Uma abordagem Consolidada** é um híbrido de Registro e Hub de Transações. Os sistemas de registro gerenciam Dados Mestres locais para seus aplicativos. Os Dados Mestres são consolidados dentro de um repositório comum e disponibilizados a partir de um hub de compartilhamento de dados, o sistema de referência para Dados Mestres. Isso elimina a necessidade de acessar diretamente dos sistemas de registro. A abordagem Consolidated fornece uma visão empresarial com impacto limitado nos sistemas de registro. No entanto, isso implica na replicação de dados e haverá latência entre o hub e os sistemas de registro.

Figura 77 Compartilhamento de dados mestre
Exemplo de Arquitetura

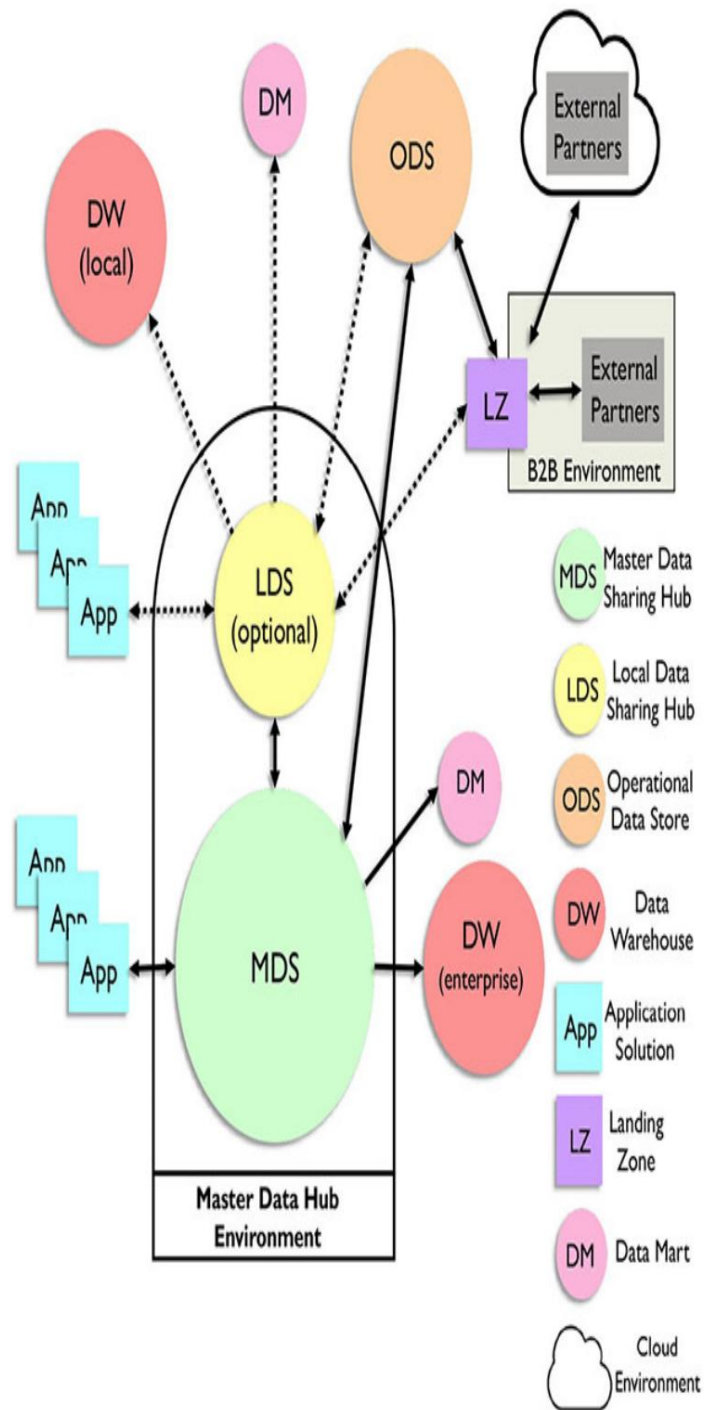


Figure 77 Master Data Sharing Architecture Example

2. Atividades

Conforme enfatizado na [Seção 1.3.1](#), Dados Mestres e Dados de Referência compartilham certas características (eles são recursos compartilhados que fornecem contexto e significado para outros dados e devem ser gerenciados no nível empresarial), mas também diferem em maneiras importantes (conjuntos de Dados de Referência são menores, menos voláteis, não exigem correspondência, mesclagem e vinculação, etc.). A seção de atividades descreverá primeiro as atividades associadas ao MDM e, em seguida, descreverá aquelas relacionadas aos Dados de Referência.

2.1 Atividades MDM

2.1.1 Definir Drivers e Requisitos de MDM Cada

organização tem diferentes drivers e obstáculos de MDM, influenciados pelo número e tipo de sistemas, sua idade, os processos de negócios que eles suportam e como os dados são usados para transações e análises. Os drivers geralmente incluem oportunidades para melhorar o atendimento ao cliente e/ou eficiência operacional, bem como para reduzir riscos relacionados à privacidade e conformidade. Os obstáculos incluem diferenças no significado e estrutura dos dados entre os sistemas. Eles geralmente estão vinculados a barreiras culturais – algumas unidades de negócios podem não querer incorrer nos custos de mudar seus processos, mesmo que a mudança seja apresentada como boa para a empresa como um todo.

É relativamente fácil definir requisitos para Master Data dentro de um aplicativo. É mais difícil definir requisitos padrão entre aplicativos. A maioria das organizações desejará abordar uma área de assunto de Master Data, ou mesmo uma entidade, por vez. Priorize os esforços de Master Data com base no custo/benefício das melhorias propostas e na complexidade relativa da área de assunto de Master Data. Comece com a categoria mais simples para aprender com o processo.

2.1.2 Avaliar e analisar fontes de dados

Dados em aplicativos existentes formam a base de um esforço de Gerenciamento de Dados Mestres. É importante entender a estrutura e o conteúdo desses dados e os processos pelos quais eles são coletados ou criados. Um resultado de um esforço de MDM pode ser melhorias em

Metadados gerados por meio do esforço para avaliar a qualidade dos dados existentes. Um objetivo da avaliação é entender o quão completos os dados são com relação aos atributos que compõem o Master Data. Esse processo inclui esclarecer as definições e a granularidade desses atributos.

Problemas semânticos surgirão em algum momento ao definir e descrever atributos. Os Data Stewards precisarão colaborar com as áreas de negócios na reconciliação e acordo sobre nomenclatura de atributos e definições de nível empresarial. (Consulte os Capítulos 3 e 13.)

A outra parte da avaliação de fontes é entender a qualidade dos dados. Problemas de qualidade de dados complicarão um projeto de Master Data, então o processo de avaliação deve incluir o tratamento das causas raiz dos problemas de dados. Nunca presuma que os dados serão de alta qualidade – é mais seguro presumir que não são de alta qualidade. Sempre avalie sua qualidade e adequação para um ambiente de Master Data.

O maior desafio, como observado, será a disparidade entre as fontes. Dados podem ser de alta qualidade dentro de qualquer fonte, mas ainda não se encaixam com dados de outras fontes, devido a diferenças estruturais e diferenças nos valores pelos quais atributos semelhantes são representados. Iniciativas de Master Data fornecem a oportunidade de definir e implementar padrões em aplicativos nos quais dados são criados ou coletados.

Para algumas entidades de Master Data, como cliente, consumidor ou fornecedor, é possível comprar dados padronizados (como Diretórios de Referência) para habilitar o esforço de MDM. Vários fornecedores têm serviços que fornecerão dados limpos relacionados a pessoas individuais ou entidades comerciais ou profissões (por exemplo, profissionais de saúde), que podem ser comparados aos dados internos de uma organização para melhorar as informações de contato, endereços e nomes (consulte o [Capítulo 10](#)). Além de avaliar a qualidade dos dados existentes, também é necessário entender a tecnologia que dá suporte à coleta de entradas para um esforço de MDM. A tecnologia existente influenciará a abordagem arquitetônica do MDM.

2.1.3 Definir abordagem arquitetônica A

abordagem arquitetônica para MDM depende da estratégia de negócios,

plataformas de fontes de dados existentes e os dados em si, particularmente sua linhagem e volatilidade, e as implicações de alta ou baixa latência.

A arquitetura deve levar em conta os modelos de consumo e compartilhamento de dados. O ferramental para manutenção depende tanto dos requisitos de negócios quanto das opções de arquitetura. O ferramental ajuda a definir e depende da abordagem de administração e manutenção.

O número de sistemas de origem a serem integrados à solução Master Data e as plataformas desses sistemas precisam ser contabilizados ao determinar a abordagem para integração. O tamanho e a distribuição geográfica de uma organização também influenciarão a abordagem de integração. Organizações pequenas podem utilizar efetivamente um hub de transações, enquanto uma organização global com vários sistemas tem mais probabilidade de utilizar um registro. Uma organização com unidades de negócios "isoladas" e vários sistemas de origem pode decidir que uma abordagem consolidada é o caminho correto a seguir. Especialistas em domínio de negócios, arquitetos de dados e arquitetos corporativos devem fornecer perspectiva sobre a abordagem.

A arquitetura do hub de compartilhamento de dados é particularmente útil quando não há um sistema claro de registro para Master Data. Nesse caso, vários sistemas fornecem dados. Novos dados ou atualizações de um sistema podem ser reconciliados com dados já fornecidos por outro sistema. O hub de compartilhamento de dados se torna a fonte do conteúdo do Master Data para data warehouses ou marts, reduzindo a complexidade das extrações e o tempo de processamento para transformação, correção e reconciliação de dados. É claro que os data warehouses devem refletir as alterações feitas no hub de compartilhamento de dados para fins históricos, enquanto o próprio hub de compartilhamento de dados pode precisar refletir apenas o estado atual.

2.1.4 Dados Mestres do Modelo

Master Data Management é um processo de integração de dados. Para atingir resultados consistentes e gerenciar a integração de novas fontes conforme uma organização se expande, é necessário modelar os dados dentro das áreas de assunto. Um modelo lógico ou canônico pode ser definido sobre as áreas de assunto dentro do hub de compartilhamento de dados. Isso permitiria o estabelecimento de definições de nível empresarial de entidades e atributos de área de assunto. (Consulte

Capítulos 5 e 8.) _

2.1.5 Definir processos de administração e manutenção

Soluções técnicas podem fazer um trabalho notável combinando, mesclando e gerenciando identificadores para registros mestres. No entanto, o processo também requer administração, não apenas para lidar com registros que saem do processo, mas também para remediar e melhorar os processos que os fazem cair em primeiro lugar. Projetos de MDM devem contabilizar os recursos necessários para dar suporte à qualidade contínua dos Dados Mestres. É necessário analisar registros, fornecer feedback aos sistemas de origem e fornecer informações que possam ser usadas para ajustar e melhorar os algoritmos que orientam a solução MDM.

2.1.6 Estabelecer Políticas de Governança para Impor o Uso de Dados

Mestres O lançamento inicial de um esforço de Dados Mestres é desafiador e exige muito foco. Os benefícios reais (eficiência operacional, maior qualidade, melhor atendimento ao cliente) vêm quando as pessoas e os sistemas começam a usar os Dados Mestres. O esforço geral tem que incluir um roteiro para que os sistemas adotem valores mestres e identificadores como entrada para os processos. Estabeleça circuitos fechados unidirecionais entre sistemas para manter a consistência de valores entre sistemas.

2.2 Atividades de Dados de Referência

2.2.1 Definir Drivers e Requisitos Os

principais drivers para o Reference Data Management são eficiência operacional e maior qualidade de dados. Gerenciar Reference Data centralmente é mais econômico do que ter várias unidades de negócios mantendo seus próprios conjuntos de dados. Também reduz o risco de inconsistência entre sistemas. Dito isso, alguns Reference Data sets são mais importantes do que outros; Reference Data sets complexos dão mais trabalho para configurar e manter do que os simples. Os Reference Data sets mais importantes devem direcionar os requisitos para um sistema Reference Data Management. Uma vez que tal sistema esteja em vigor, novos Reference Data sets podem ser configurados como parte dos projetos. Reference Data sets existentes devem ser mantidos com base em um cronograma publicado.

2.2.2 Avaliar fontes de dados

A maioria dos conjuntos de Dados de Referência padrão da indústria pode ser obtida das organizações que os criam e mantêm. Algumas organizações fornecem esses dados gratuitamente. Outras cobram uma taxa. Os intermediários também empacotam e vendem Dados de Referência, geralmente com recursos de valor agregado. Dependendo do número e do tipo de conjuntos de Dados de Referência necessários para uma organização, pode ser melhor comprar de um fornecedor, especialmente se esse fornecedor garantir a entrega de atualizações em um cronograma definido e executar o controle de qualidade básico nos dados.

A maioria das organizações também depende de Dados de Referência que são criados e mantidos internamente. Determinar a fonte para dados de referência internos ou locais é frequentemente mais desafiador do que fazer isso para Dados de Referência padrão do setor. Como é o caso com Dados Mestres, fontes internas para Dados de Referência devem ser identificadas, comparadas e avaliadas. Proprietários de dados existentes devem entender os benefícios do gerenciamento central e concordar em dar suporte a processos para administrar os dados para o bem da empresa.

2.2.3 Definir Abordagem Arquitetural Antes

de comprar ou construir uma ferramenta para gerenciar Dados de Referência, é essencial levar em conta os requisitos e os desafios impostos pelos Dados de Referência a serem gerenciados. Por exemplo, a volatilidade dos dados (a maioria dos Dados de Referência é relativamente estática, mas alguns são bastante voláteis), a frequência de atualizações e os modelos de consumo. Determine se é necessário manter dados históricos sobre alterações nos valores ou nas definições dos valores. Se a organização comprar dados de um fornecedor, leve em conta o método de entrega e integração.

A abordagem arquitetônica precisa reconhecer que, invariavelmente, alguns Dados de Referência precisarão ser atualizados manualmente. Garanta que a interface para atualizações seja direta e possa ser configurada para impor regras básicas de entrada de dados, como garantir que relacionamentos pai/filho sejam mantidos em Dados de Referência que incluam hierarquias. A ferramenta RDM deve permitir que os Stewards façam atualizações ad hoc sem a necessidade de suporte técnico e deve incluir

fluxos de trabalho para garantir que aprovações e notificações sejam automatizadas. Os administradores de dados devem agendar atualizações conhecidas para se alinharem com a publicação de novos códigos. Os consumidores de dados devem ser informados de todas as alterações. Em casos em que os dados de referência direcionam a lógica de programação, o impacto potencial das alterações deve ser avaliado e contabilizado antes que as alterações sejam introduzidas.

2.2.4 Conjuntos de Dados de Referência de

Modelo Muitas pessoas pensam em Dados de Referência como simplesmente códigos e descrições. No entanto, muitos Dados de Referência são mais complicados do que isso. Por exemplo, um conjunto de dados de CEP geralmente incluirá informações sobre estado e condado, bem como outros atributos geopolíticos. Para fins de permitir o uso a longo prazo e estabelecer Metadados precisos, bem como para o próprio processo de manutenção, é valioso criar modelos de dados de conjuntos de Dados de Referência. Os modelos ajudam os consumidores de dados a entender os relacionamentos dentro do conjunto de Dados de Referência e podem ser usados para estabelecer regras de qualidade de dados.

2.2.5 Definir Processos de Administração e Manutenção Dados de

Referência requerem administração para garantir que os valores estejam completos e atualizados e que as definições sejam claras e compreensíveis. Em alguns casos, os administradores serão diretamente responsáveis pela manutenção prática dos Dados de Referência; em outros casos, eles podem facilitar o processo. Por exemplo, se várias unidades de negócios diferentes exigirem Dados de Referência para dar suporte ao mesmo conceito, um administrador pode facilitar discussões que definam valores comuns em uma faixa de pedestres.

Como parte do processo de administração, é útil capturar Metadados básicos sobre cada conjunto de Dados de Referência. Isso pode incluir: nome do administrador, organização de origem, frequência esperada de atualizações, cronograma para atualizações, processos usando os Dados de Referência, se versões históricas dos dados precisam ser retidas e mais (consulte [a Seção 1.3.2.6](#)). Documentar quais processos usam Dados de Referência permitirá uma comunicação mais eficaz sobre alterações nos dados.

Muitas ferramentas de gerenciamento de dados de referência incluem fluxos de trabalho para

gerenciar revisão e aprovação de alterações em Dados de Referência. Esses fluxos de trabalho em si dependem da identificação de quem dentro de uma organização é responsável pelo conteúdo de Dados de Referência.

2.2.6 Estabelecer Políticas de Governança de Dados de Referência

Uma organização só obtém valor de um repositório de Dados de Referência gerenciado centralmente se as pessoas realmente usarem os dados daquele repositório. É importante ter políticas em vigor que governem a qualidade e exijam o uso de Dados de Referência daquele repositório, seja diretamente por meio da publicação daquele repositório ou indiretamente de um sistema de referência que é preenchido com dados do repositório central.

3. Ferramentas e Técnicas

O MDM requer ferramentas projetadas especificamente para habilitar o gerenciamento de identidade. O Master Data Management pode ser implementado por meio de ferramentas de integração de dados, ferramentas de remediação de dados, armazenamentos de dados operacionais (ODS), hubs de compartilhamento de dados (DSH) ou aplicativos MDM especializados. Vários fornecedores oferecem soluções que podem cobrir uma ou mais áreas de assunto de Master Data. Outros fornecedores promovem o uso de seus produtos de software de integração de dados e serviços de implementação para criar soluções de Master Data personalizadas.

Soluções empacotadas para produtos, contas e partes, bem como serviços de verificação de qualidade de dados empacotados, podem dar início a grandes programas.

A incorporação desses serviços pode permitir que as organizações usem as melhores soluções, ao mesmo tempo em que as integram à sua arquitetura geral de negócios para atender a necessidades específicas.

4. Diretrizes de implementação

Master e Reference Data Management são formas de integração de dados. Os princípios de implementação que se aplicam à integração de dados e à interoperabilidade se aplicam ao MDM e ao RDM. (Consulte [o Capítulo 8](#).)

Os recursos de MDM e RDM não podem ser implementados da noite para o dia.

Soluções exigem conhecimento técnico e comercial especializado.

As organizações devem esperar implementar soluções de Dados Mestres e de Referência de forma incremental por meio de uma série de projetos definidos em um roteiro de implementação, priorizados com base nas necessidades do negócio e orientados por uma arquitetura geral.

Observe que os programas de MDM falharão sem governança adequada. Os profissionais de governança de dados devem entender os desafios de MDM e RDM e avaliar a maturidade e a capacidade da organização de enfrentá-los. (Consulte [o Capítulo 15](#).)

4.1 Aderir à Arquitetura de Dados Mestres

Estabelecer e seguir a arquitetura de referência adequada é essencial para gerenciar e compartilhar Master Data em uma organização. A abordagem de integração deve levar em conta a estrutura organizacional do negócio, o número de sistemas distintos de registro, a implementação da governança de dados, a importância do acesso e latência dos valores de dados e o número de sistemas e aplicativos consumidores.

4.2 Monitorar a movimentação de dados

Os processos de integração de dados para Dados Mestres e de Referência devem ser projetados para garantir a extração e distribuição oportunas de dados por toda a organização. À medida que os dados fluem dentro de um ambiente de compartilhamento de Dados Mestres ou de Referência, o fluxo de dados deve ser monitorado para:

- Mostre como os dados são compartilhados e usados em toda a organização
- Identificar a linhagem de dados de/para sistemas e aplicativos administrativos
- Auxiliar na análise da causa raiz dos problemas
- Demonstrar eficácia das técnicas de ingestão de dados e integração de consumo
- Indica a latência dos valores de dados dos sistemas de origem até o consumo

- Determinar a validade das regras de negócios e transformações executadas nos componentes de integração

4.3 Gerenciar Mudança de Dados de Referência Como

Dados de Referência são um recurso compartilhado, eles não podem ser alterados arbitrariamente. A chave para o Gerenciamento de Dados de Referência bem-sucedido é a disposição organizacional de abrir mão do controle local de dados compartilhados. Para sustentar esse suporte, forneça canais para receber e responder a solicitações de alterações em Dados de Referência. O Data Governance Council deve garantir que políticas e procedimentos sejam implementados para lidar com alterações em dados dentro de ambientes de Dados Mestres e de referência.

Alterações nos Dados de Referência precisarão ser gerenciadas. Pequenas alterações podem afetar algumas linhas de dados. Por exemplo, quando a União Soviética se dividiu em estados independentes, o termo *União Soviética* foi descontinuado e novos códigos foram adicionados. No setor de saúde, os códigos de procedimento e diagnóstico são atualizados anualmente para levar em conta o refinamento dos códigos existentes, a obsolescência dos códigos e a introdução de novos códigos. Grandes revisões nos Dados de Referência impactam a estrutura dos dados. Por exemplo, os Códigos de Diagnóstico da CID-10 são estruturados de maneiras muito diferentes da CID-9. A CID10 tem um formato diferente. Existem valores diferentes para os mesmos conceitos. Mais importante, a CID-10 tem princípios adicionais de organização. Os códigos da CID10 têm uma granularidade diferente e são muito mais específicos, então mais informações são transmitidas em um único código. Consequentemente, há muito mais deles (em 2015, havia 68.000 códigos CID-10, em comparação com 13.000 CID-9s).⁶² O uso obrigatório de códigos CID-10 nos EUA em 2015 exigiu um planejamento significativo. As empresas de --- saúde precisaram fazer mudanças no sistema, bem como ajustes nos relatórios impactados para levar em conta o novo padrão.

Os tipos de alterações incluem:

- Alterações no nível da linha para conjuntos de dados de referência externos
- Alterações estruturais em conjuntos de dados de referência externos
- Alterações no nível da linha para conjuntos de dados de referência internos
- Alterações estruturais nos conjuntos de dados de referência internos
- Criação de novos conjuntos de dados de referência

As mudanças podem ser planejadas/programadas ou ad hoc. Mudanças planejadas, como atualizações mensais ou anuais para códigos padrão da indústria, exigem menos governança do que atualizações ad hoc. O processo para solicitar novos conjuntos de Dados de Referência deve levar em conta usos potenciais além daqueles do solicitante original.

As solicitações de mudança devem seguir um processo definido, conforme ilustrado na Figura 78. Quando as solicitações são recebidas, as partes interessadas devem ser notificadas para que os impactos possam ser avaliados. Se as mudanças precisarem de aprovação, discussões devem ser realizadas para obter essa aprovação. As mudanças devem ser comunicadas.

Figura 78 Dados de referência Processo de solicitação de mudança



Figure 78 Reference Data Change Request Process

4.4 Acordos de compartilhamento de dados Compartilhar

e usar dados de referência e mestres em uma organização requer colaboração entre várias partes internas à organização.

organização e, às vezes, com partes externas a ela. Para garantir acesso e uso adequados, estabeleça acordos de compartilhamento que estipulem quais dados podem ser compartilhados e sob quais condições. Ter esses acordos em vigor ajudará quando surgirem problemas relacionados à disponibilidade de dados dentro ou à qualidade dos dados trazidos para o ambiente de compartilhamento de dados. Esse esforço deve ser conduzido pelo programa de Governança de Dados. Ele pode envolver Arquitetos de Dados, Provedores de Dados, Administradores de Dados, Desenvolvedores de Aplicativos, Analistas de Negócios, bem como Oficiais de Conformidade/Privacidade e Oficiais de Segurança.

Os responsáveis pelo ambiente de compartilhamento de dados têm a obrigação de fornecer dados de alta qualidade aos consumidores de dados downstream. Para cumprir essa responsabilidade, eles dependem de sistemas upstream. SLAs e métricas devem ser estabelecidos para medir a disponibilidade e a qualidade dos dados compartilhados. Processos devem ser colocados em prática para abordar as causas raiz dos problemas com a qualidade ou disponibilidade dos dados. Uma abordagem padrão para comunicações deve ser colocada em prática para manter todas as partes afetadas informadas sobre a existência de problemas e o status dos esforços de correção.

(Consulte [o Capítulo 8](#).)

5. Organização e Mudança Cultural

O Reference and Master Data Management exige que as pessoas abram mão do controle de alguns de seus dados e processos para criar recursos compartilhados. Nem sempre é fácil fazer isso. Embora os profissionais de gerenciamento de dados possam ver que dados gerenciados localmente são arriscados, as pessoas que os gerenciam localmente precisam fazer seu trabalho e podem perceber os esforços de MDM ou RDM como uma adição de complicação aos seus processos.

Felizmente, a maioria das pessoas reconhece que esses esforços fazem sentido fundamental. É melhor ter uma visão precisa e completa de um único cliente do que ter múltiplas visões parciais.

Melhorar a disponibilidade e a qualidade dos dados de referência e mestres exigirá, sem dúvida, mudanças nos procedimentos e práticas tradicionais. As soluções devem ser delimitadas e implementadas com base na prontidão organizacional atual e nas necessidades futuras vinculadas à

missão e visão da organização.

Talvez a mudança cultural mais desafiadora seja central para a governança: determinar quais indivíduos são responsáveis por quais decisões — administradores de dados de negócios, arquitetos, gerentes e executivos — e quais decisões as equipes de administração de dados, comitês de direção de programas e o Conselho de Governança de Dados devem tomar de forma colaborativa.

6. Governança de dados mestres e de referência

Como são recursos compartilhados, os Dados Mestres e de Referência exigem governança e administração. Nem todas as inconsistências de dados podem ser resolvidas por meio da automação. Algumas exigem que as pessoas conversem entre si. Sem governança, as soluções de Dados Mestres e de Referência serão apenas utilitários adicionais de integração de dados, incapazes de entregar todo o seu potencial.

Os processos de governança determinarão:

- As fontes de dados a serem integradas
- As regras de qualidade de dados a serem aplicadas
- As condições de uso regras a serem seguidas
- As atividades a serem monitoradas e a frequência do monitoramento
- Os níveis de prioridade e resposta dos esforços de administração de dados
- Como a informação deve ser representada para atender às necessidades das partes interessadas
- Portões de aprovação padrão, expectativas na implantação de RDM e MDM

Os processos de governança também reúnem as partes interessadas em conformidade e legais com os consumidores de informações para garantir os riscos organizacionais

são mitigados por meio da definição e incorporação de políticas de privacidade, segurança e retenção.

Como um processo contínuo, a governança de dados deve ter a capacidade de revisar, receber e considerar novos requisitos e mudanças nas regras existentes, ao mesmo tempo em que disponibiliza princípios, regras e diretrizes para aqueles que usam Dados de Referência e Dados Mestres.

6.1 Métricas

Certas métricas podem ser vinculadas à qualidade dos Dados Mestres e de Referência e aos processos que dão suporte a esses esforços:

- **Qualidade e conformidade dos dados:** os painéis do DQ podem descrever a qualidade dos Dados de Referência e Mestres. Essas métricas devem denotar a confiança (como uma porcentagem) de uma entidade de área de assunto ou atributo associado e sua adequação à finalidade para uso em toda a organização.
- **Atividade de alteração de dados:** Auditar a linhagem de dados confiáveis é essencial para melhorar a qualidade dos dados em um ambiente de compartilhamento de dados. Métricas devem denotar a taxa de alteração dos valores de dados. Essas métricas fornecerão insights aos sistemas que fornecem dados ao ambiente de compartilhamento e podem ser usadas para ajustar algoritmos em processos de MDM.
- **Ingestão e consumo de dados:** os dados são fornecidos por sistemas upstream e usados por sistemas e processos downstream. Essas métricas devem denotar e rastrear quais sistemas estão contribuindo com dados e quais áreas de negócios estão assinando dados do ambiente de compartilhamento.
- **Acordos de Nível de Serviço:** SLAs devem ser estabelecidos e comunicados aos colaboradores e assinantes para garantir o uso e a adoção do ambiente de compartilhamento de dados. O nível de adesão aos SLAs pode fornecer insights sobre os processos de suporte e os problemas técnicos e de dados que podem retardar a aplicação do MDM.
- **Cobertura do Data Steward:** Essas métricas devem observar o

nome ou grupo responsável pelo conteúdo dos dados e com que frequência a cobertura é avaliada. Eles podem ser usados para identificar lacunas no suporte.

- **Custo Total de Propriedade:** Existem múltiplos fatores dessa métrica e diferentes maneiras de representá-la. De uma visão de solução, os custos podem incluir infraestrutura de ambiente, licenças de software, equipe de suporte, taxas de consultoria, treinamento, etc. A eficácia dessa métrica é amplamente baseada em sua aplicação consistente em toda a organização.
- **Volume e uso de compartilhamento de dados:** os volumes de ingestão e consumo de dados precisam ser rastreados para determinar a eficácia do ambiente de compartilhamento de dados. Essas métricas devem denotar o volume e a velocidade dos dados definidos, ingeridos e assinados para e do ambiente de compartilhamento de dados.

7. Trabalhos Citados / Recomendados

Abbas, junho. *Estruturas para organizar o conhecimento: explorando taxonomias, ontologias e outros esquemas*. Neal-Schuman Publishers, 2010. Impresso.

Abernethy, Kenneth e J. Thomas Allen. *Explorando o Domínio Digital: Uma Introdução a Computadores e Fluência de Informação*. 2ª ed., 2004. Impresso.

Allen Mark e Dalton Cervo. *Gerenciamento de dados mestres multidomínio: MDM avançado e governança de dados na prática*. Morgan Kaufmann, 2015. Impresso.

Bean, James. *XML para Arquitetos de Dados: Projetando para Reutilização e Integração*. Morgan Kaufmann, 2003. Impresso. A Série Morgan Kaufmann em Sistemas de Gerenciamento de Dados.

Berson, Alex e Larry Dubov. *Master Data Management e Customer Data Integration para uma empresa global*. McGraw-Hill, 2007. Impresso.