



Enap

# Governança de Dados

Módulo

3

Gestão inteligente de Dados



## **Fundação Escola Nacional de Administração Pública**

### **Presidente**

Diogo Godinho Ramos Costa

### **Diretor de Educação Continuada**

Paulo Marques

### **Coordenador-Geral de Educação a Distância**

Carlos Eduardo dos Santos

### **Conteudista/s**

Wellington Luiz Barbosa (conteudista, 2019)

Roberto Shayer Lyra (conteudista, 2019)

**Curso produzido em Brasília 2021.**



Enap, 2021

### **Enap Escola Nacional de Administração Pública**

Diretoria de Educação Continuada

SAIS - Área 2-A - 70610-900 — Brasília, DF



# Sumário

<b>1. Governança de Dados</b> .....	<b>5</b>
1.1 Governança de dados como supervisão.....	5
1.2 Direcionadores (Drivers) de Negócios.....	9
1.3 - Características do Programa de Controle de Dados.....	11
1.4 Modelos de Governança de Dados.....	13
1.5 Gerenciamento de dados.....	16
1.6 Iniciando o Controle de Dados.....	16
1.7 Governança de Dados Sustentável.....	18
<b>2. Planejamento e design no gerenciamento do ciclo de vida dos dados</b> .....	<b>18</b>
2.1 Arquitetura de Negócio.....	19
2.2 O Quadro Zachman.....	20
2.3 Arquitetura de Dados.....	22
2.4 Artefatos de arquitetura de dados.....	23
2.5 Modelagem de dados.....	30
<b>3. Habilitando e mantendo dados</b> .....	<b>35</b>
3.1 Armazenamento e operações de dados.....	35
3.2 Integração e interoperabilidade de dados.....	36
3.3 Armazenamento de Dados.....	43
3.4 Gerenciamento de Dados de referência.....	46
3.5 Gerenciamento de Dados Mestre.....	47
3.6 Gerenciamento de documentos e conteúdo.....	48
3.7 <i>Big Data</i> .....	50
<b>4. Usando e aprimorando dados</b> .....	<b>52</b>
4.1 Inteligência do Negócio.....	52
4.2 Ciência de Dados.....	53
4.3 Análise preditiva e prescritiva.....	53
4.4 Visualização de dados.....	55
4.5 Monetização de Dados.....	56



<b>5. Proteção de Dados, Privacidade, Segurança e Gerenciamento de Riscos</b> .....	<b>57</b>
5.1 Objetivos de segurança de dados .....	58
5.2 Princípios de segurança de dados .....	60
5.3. Segurança de dados e gerenciamento de dados corporativos .....	61
5.4 Metadados de Segurança de Dados .....	63
5.5 Arquitetura de segurança de dados .....	67
5.6 Planejando a segurança de dados .....	68
<b>Referências</b> .....	<b>70</b>



## Módulo

# 3

## Gestão inteligente de Dados

### 1. Governança de Dados

A governança de dados (GD) é definida como o exercício de autoridade sobre o gerenciamento de dados, promovendo a redução dos riscos associados e orientando o uso estratégico dos dados.

As organizações que estabelecem um programa formal de governança de dados exercem autoridade e controle com maior impacto e consistência. Com isso, são mais propensas a alavancar estrategicamente o valor dos seus ativos de dados.

#### 1.1 Governança de dados como supervisão

A função de controle de dados orienta todas as outras funções de gerenciamento de dados. O objetivo do controle de dados é garantir que os dados sejam gerenciados corretamente, de acordo com políticas e práticas recomendadas. Uma analogia comum é equiparar governança de dados às atividades de auditoria. Auditores e controladores definem as regras para o gerenciamento de ativos financeiros. Os profissionais de governança de dados definem regras para gerenciar ativos de dados. Outras áreas cumprem essas regras. Em ambos os casos, a governança de dados não é única; essa função de supervisão deve ser mantida depois de estabelecida. Os princípios para governança de dados precisam ser incorporados ao ciclo de vida de gerenciamento de dados e às atividades fundamentais. Um programa de governança de dados requer compromisso contínuo para garantir que uma organização obtenha valor de seus dados e promova a redução dos riscos associados.

A governança de dados se concentra em como as decisões são tomadas sobre os dados e como as pessoas e os processos serão impactados. O escopo e o objetivo de um programa de governança de dados incluem:

- **Supervisão:** Garantir que todas as áreas funcionais de governança de dados sigam os princípios orientadores da organização.
- **Estratégia:** definir, comunicar e direcionar a execução da estratégia.
- **Política:** Definir e aplicar as políticas relacionadas ao gerenciamento, acesso, uso, segurança e qualidade de dados e metadados.
- **Padrões e qualidade:** definir e aplicar padrões de qualidade e arquitetura de dados.

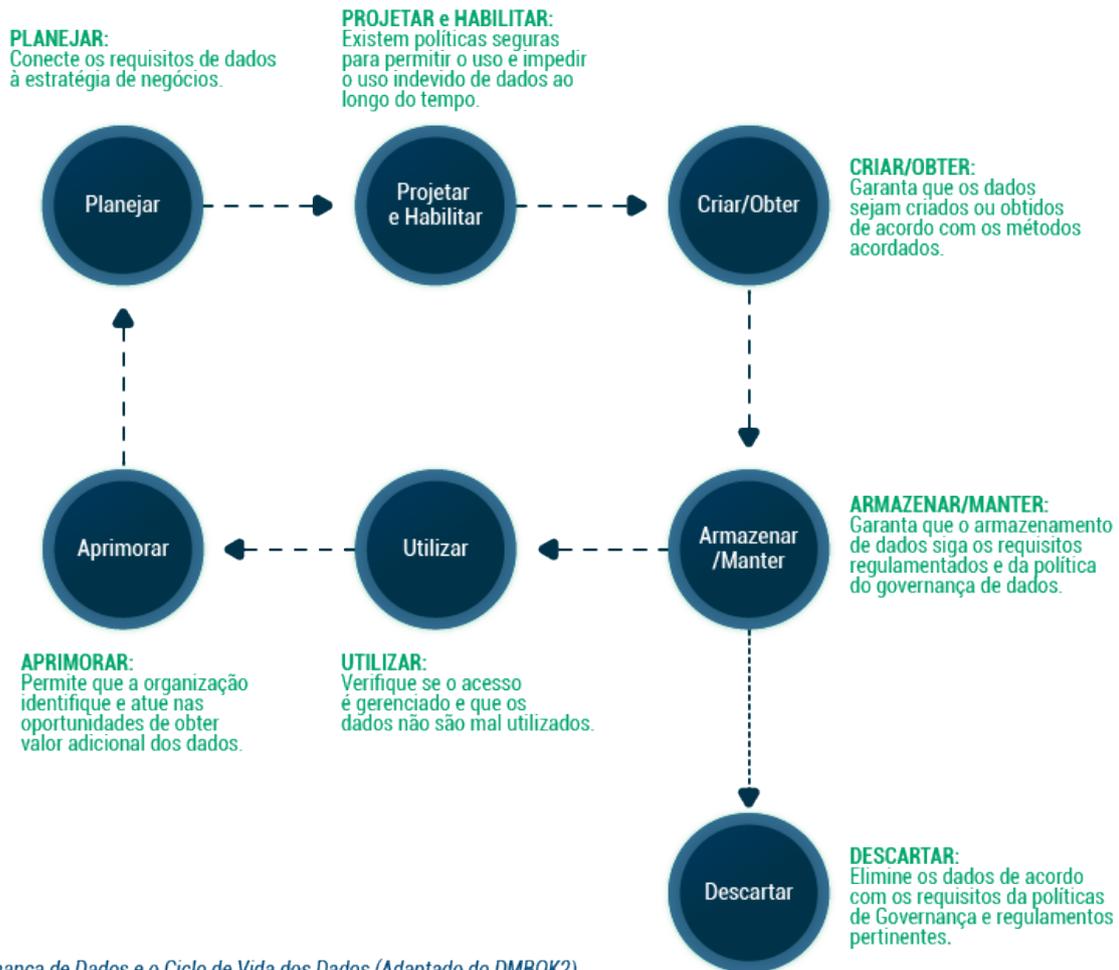
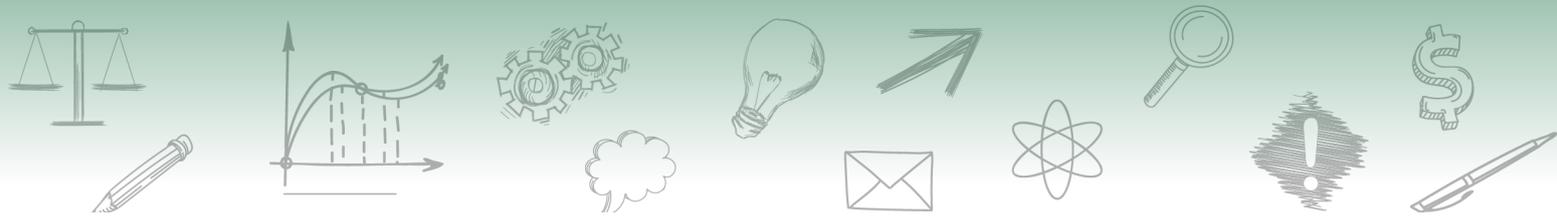


- **Gestão:** Fornecer insumos de controle e acompanhamento das áreas de qualidade e gerenciamento de dados.
- **Conformidade:** garantir que a organização possa atender aos requisitos de conformidade regulamentar relacionados a dados.
- **Gerenciamento de problemas:** identificar, definir, escalar e resolver problemas relacionados à segurança, acesso, qualidade, conformidade regulamentar, propriedade de dados, política, padrões, terminologia ou procedimentos de governança de dados.
- **Projetos de gerenciamento de dados:** patrocinar esforços para aprimorar as práticas de gerenciamento de dados.
- **Avaliação de ativos de dados:** definir padrões e processos consistentemente ao valor estratégico dos ativos de dados.



*Governança de Dados/Gerenciamento de Dados (DMBOK2)*

Um programa de governança de dados irá estabelecer princípios, políticas e procedimentos para se implantar uma estrutura de governança de dados, bem como as ações de mudanças de comportamento necessárias para o gerenciamento de dados no decorrer do ciclo de vida dos dados.



Governança de Dados e o Ciclo de Vida dos Dados (Adaptado do DMBOK2)

O programa de governança é planejado com base no resultado de um modelo de maturidade de boas práticas. Isso implica no gerenciamento de mudanças, bem como na necessidade de patrocínio da alta liderança, como Chief Risk Officer e/ou Chief Data Officer. Ambos poderão atuar como Curadores de Dados ou Data Steward - papel fundamental para as atividades de governança. São essas atividades que impulsionam o processo de qualidade de dados e de informações, permitindo fazer a gestão dos dados de forma aderente as necessidades de negócio.

## SAIBA MAIS

### Chief Risk Officer (CRO) or Chief Risk Management Officer (CRMO)

O Diretor de Riscos (CRO) ou o Diretor de Gerenciamento de Riscos (CRMO) é o executivo responsável por permitir a governança eficiente e eficaz de riscos significativos e oportunidades relacionadas a uma organização e seus vários segmentos. Os riscos são geralmente classificados como estratégicos, de reputação, operacionais, financeiros ou relacionados à conformidade. Os



CROs são responsáveis perante Comitês e/ou Conselhos por permitir que os negócios equilibrem risco e resultados.

Em organizações mais complexas, eles geralmente são responsáveis por coordenar o Gerenciamento de Riscos Corporativos da organização (ERM). O CRO é responsável por avaliar e mitigar ameaças competitivas, regulatórias e tecnológicas significativas aos resultados de organização. As funções e responsabilidades de CRO variam de acordo com o tamanho da organização e do setor. O CRO trabalha para garantir que a organização esteja em conformidade com os regulamentos governamentais, como *Sarbanes-Oxley*, e analisa fatores que podem afetar negativamente os investimentos. Normalmente, o CRO é responsável pelas operações de gerenciamento de riscos da organização, incluindo gerenciamento, identificação, avaliação, relatório e supervisão dos riscos da organização, externa e internamente, para a organização e trabalha diligentemente com a gerência sênior, como CEO e CFO.

O papel do Chief Risk Officer (CRO) está se tornando cada vez mais relevante. De acordo com Watson, a maioria dos CROs concorda que ter apenas habilidade analítica excepcional não é suficiente. Os CROs mais bem-sucedidos são capazes de combinar essas habilidades com as habilidades de negócio, estratégicas, de liderança e de comunicação para conduzir mudanças.

James Lam, um notável profissional de risco, é creditado como a primeira pessoa a cunhar o termo. Lam é a primeira pessoa a ocupar esse cargo na GE Capital em 1993. O cargo se tornou mais comum após o Acordo de Basileia, a Lei *Sarbanes-Oxley*, o *Turnbull Report*.

### O Diretor de Dados

A maioria das organizações reconhece em algum nível que os dados são um ativo corporativo. Na última década, alguns nomearam *Chief Data Officers* (CDO) para ajudar a preencher a lacuna entre tecnologia e negócios e disseminar uma estratégia de gerenciamento de dados em toda a organização em nível sênior. Este papel está em ascensão. A revista Forbes informou em janeiro de 2018 que mais de 60% das organizações da Fortune 1000 têm um CDO.

Embora os requisitos e funções de um CDO sejam específicos da cultura, estrutura organizacional e necessidades de negócios de cada organização, muitos atuam como uma combinação de estrategista de negócios, consultor, administrador de qualidade e de gerenciamento de dados.

Em 2014, a *Dataversity* publicou uma pesquisa descrevendo competências comuns para um CDO:



- Estabelecer uma estratégia de dados organizacionais.
- Alinhar requisitos centrados em dados com recursos de negócios e TI disponíveis.
- Estabelecer padrões, políticas e procedimentos de governança de dados.
- Fornecer consultoria (e talvez serviços) aos negócios para iniciativas dependentes de dados, como análise de negócios, Big Data, qualidade de dados e tecnologias de dados.
- Disseminar a importância de bons princípios de gerenciamento de informações para as partes interessadas internas e externas da organização.
- Fornecer supervisão do uso de dados em análises e inteligência de negócios.

Em um modelo operacional mais descentralizado, o CDO é responsável pela estratégia de dados.

## 1.2 Direcionadores (Drivers) de Negócios

O que é mais comum para a governança de dados é ser aderente a conformidade. Principalmente em setores fortemente regulamentados, como serviços financeiros, de saúde, de transportes, etc. Atender a uma legislação sempre em evolução requer processos aprimorados de governança de dados. O avanço da ciência de dados alavancou a necessidade de se implementar estruturas de governança de dados.

Embora a conformidade e/ou análise de dados possam impulsionar a implementação de governança, as organizações avaliam a governança de dados apenas do ponto de vista de um programa de gerenciamento de informações direcionado por outras necessidades de negócios, como o *Master Data Management* (MDM).

A governança de dados deve estar alinhada à estratégia de negócio. Os direcionadores (*drivers*) para governança de dados geralmente se concentram em:

- **Reduzir riscos**, como os relacionados à conformidade, a segurança e a privacidade dos dados.
- **Melhorar processos**, capacidade instalada de cumprir com a regulamentação, gerenciar serviços e operar com eficiência.



## SAIBA MAIS

Em termo de computação, Master Data Management (MDM) compreende um conjunto de processos e ferramentas que definem consistentemente e gerenciam os dados mestres (i.e. não transacionais) de uma organização (que pode conter dados referenciais). O MDM tem o objetivo de prover processos para coletar, agregar, combinar, consolidar, garantir a qualidade, persistir e distribuir estes dados a toda a organização para garantir a consistência e o controle nas futuras manutenções e uso desta informação.

O termo remete ao conceito de arquivo mestre de uma antiga era da computação. MDM seria o mesmo que *virtual or federated database management*.

## DESTAQUE

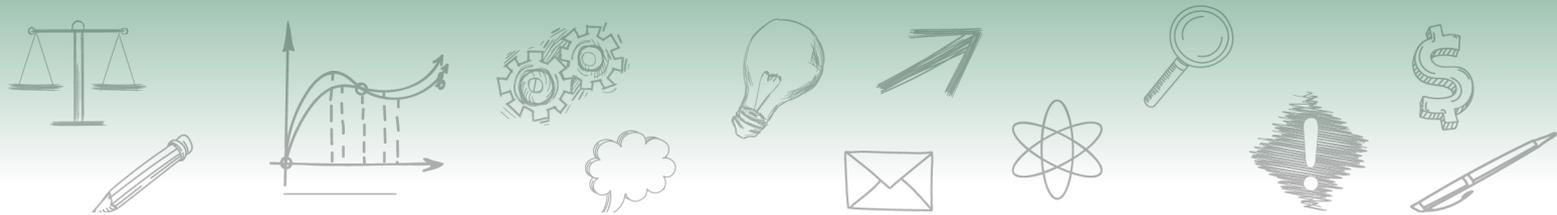
### Cadastro Base do Cidadão (CBC)

O cidadão, ao longo da vida cível, precisa realizar seu cadastro em diversos órgãos públicos para emitir um documento de identificação, como por exemplo: certidão de nascimento, RG, CPF, título de eleitor, carteira de habilitação etc. Apesar de cada documento desses ter um propósito específico, todos eles tratam de dados biográficos do cidadão, tais como: nome, data de nascimento, sexo etc.

Apesar de termos tido grandes avanços tecnológicos na prestação desses serviços públicos, eles ocorreram de maneira individualizada. O resultado negativo disso é que, além do custo do cidadão de ter de se apresentar diversas vezes ao Estado, temos seu cadastro replicado em diversas bases de dados do governo. Nesse contexto, é difícil identificar o mesmo cidadão nestas diversas bases, o que inviabiliza a integração entre os sistemas governamentais e a composição de sua trajetória pelos diversos serviços estatais. Além disso, a replicação permite que os dados estejam diferentes entre as bases de dados e não se sabe qual é o registro mais atualizado ou correto, o que gera inconsistências e perda de confiabilidade nestes dados.

O Cadastro Base do Cidadão - CBC visa unificar e melhorar as informações sobre o cidadão dentro do governo. Tem como base os cadastros hoje existentes nas bases do governo, principalmente no Cadastro de Pessoa Física mantido pela Secretaria de Receita Federal. O CBC não irá substituir nem eliminar os cadastros existentes, mas alinhar suas informações.

O CBC foi instituído pelo Decreto 10.046, de 09 de outubro de 2019, e tem como finalidade, segundo o art. 16:



**I. aprimorar a gestão de políticas públicas;**

**II. aumentar a confiabilidade dos cadastros de cidadãos existentes na administração pública, por meio de mecanismos de manutenção da integridade das bases de dados para torná-las qualificadas e consistentes;**

**III. viabilizar a criação de meio unificado de identificação do cidadão para a prestação de serviços públicos;**

**IV. disponibilizar uma interface unificada de atualização cadastral, suportada por soluções tecnológicas interoperáveis das entidades e órgãos públicos participantes do cadastro;**

**V. facilitar o compartilhamento de dados cadastrais do cidadão entre os órgãos da administração pública; e**

**VI. realizar o cruzamento de informações das bases de dados cadastrais oficiais a partir do número de inscrição do cidadão no CPF.**

#### Gestão do CBC

O Comitê Central de Governança de Dados é o colegiado responsável pela gestão do Cadastro Base do Cidadão, por disposição do mesmo decreto

#### Benefícios:

Os usuários de serviços públicos terão processos e procedimentos de atendimento simplificados e confiáveis, além da eliminação de formalidades e exigências de comprovação de identidade.

O Estado será beneficiado pela aplicação de solução tecnológica que simplificará processos e procedimentos de atendimento aos usuários dos serviços públicos e proporcionará melhores condições para o compartilhamento das informações entre órgãos de governo, aumento da segurança da informação em função da garantia de sua autenticidade;

Toda sociedade será beneficiada pelo uso de meio unificado e confiável de identificação unívoca e inequívoca de pessoas gerando confiabilidade entre Estado e sociedade.

## 1.3 Características do Programa de Controle de Dados

A governança de dados fornece os princípios, políticas, processos e indicadores necessários para orientar as atividades de gerenciamento de dados. Para atingir esse objetivo, um programa de governança de dados deve ser:



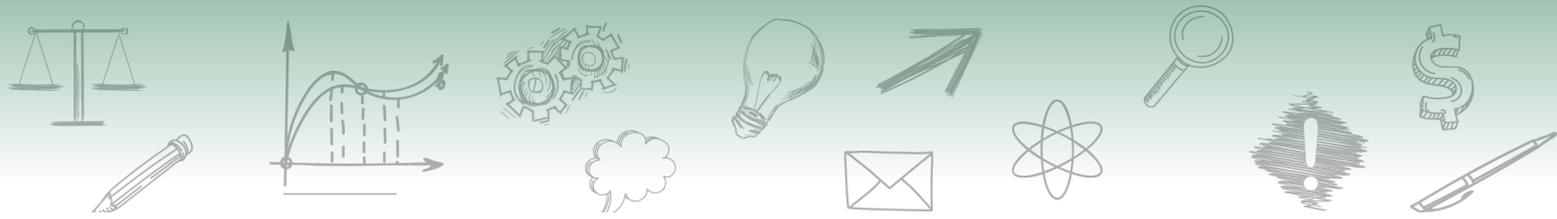
- **Sustentável:** a governança de dados é um processo contínuo que requer compromisso da organização. Exige mudanças na maneira como os dados são gerenciados e usados. Isso significa gerenciar as mudanças, além da implementação inicial de qualquer componente de controle de dados.
- **Incorporado:** a governança de dados não é um processo complementar. As atividades de governança precisam ser incorporadas as principais atividades de gestão de dados.
- **Mensurável:** a governança de dados bem implementada gera impactos positivos, mas demonstrar esse impacto requer o planejamento e definição de indicadores que represente o quanto a GD gera resultados.

A implementação de um programa de governança de dados requer compromisso com a mudança. Os seguintes princípios, desenvolvidos desde o início dos anos 2000, podem auxiliar a estabelecer uma base sólida para essa implementação:

- **Liderança e estratégia:** a governança de dados bem-sucedida começa com uma liderança visionária, comprometida e alinhada à estratégia de negócios.
- **Orientado aos negócios:** a governança de dados é um programa de negócios que deve governar as decisões de TIC relacionadas aos dados e suas interações com as áreas de negócios.
- **Responsabilidade compartilhada:** responsabilidade compartilhada entre gestores de negócios e profissionais de gerenciamento de dados.
- **Multicamada:** atua da alta administração aos níveis intermediários.
- **Baseado em estrutura:** o programa de governança de dados deve estabelecer uma estrutura que defina responsabilidades e interações.
- **Baseado em princípios:** os princípios e as políticas são a base das atividades de governança de dados.

A governança de dados pode ser entendida em termos de uma estrutura política:

- **Funções legislativas (*Legislative-like functions*):** definição de políticas, padrões e arquitetura de dados corporativos.
- **Funções judiciais (*Judicial-like functions*):** gerenciamento e escalonamento de problemas.
- **Funções executivas (*Executive functions*):** proteger e servir, responsabilidades administrativas.

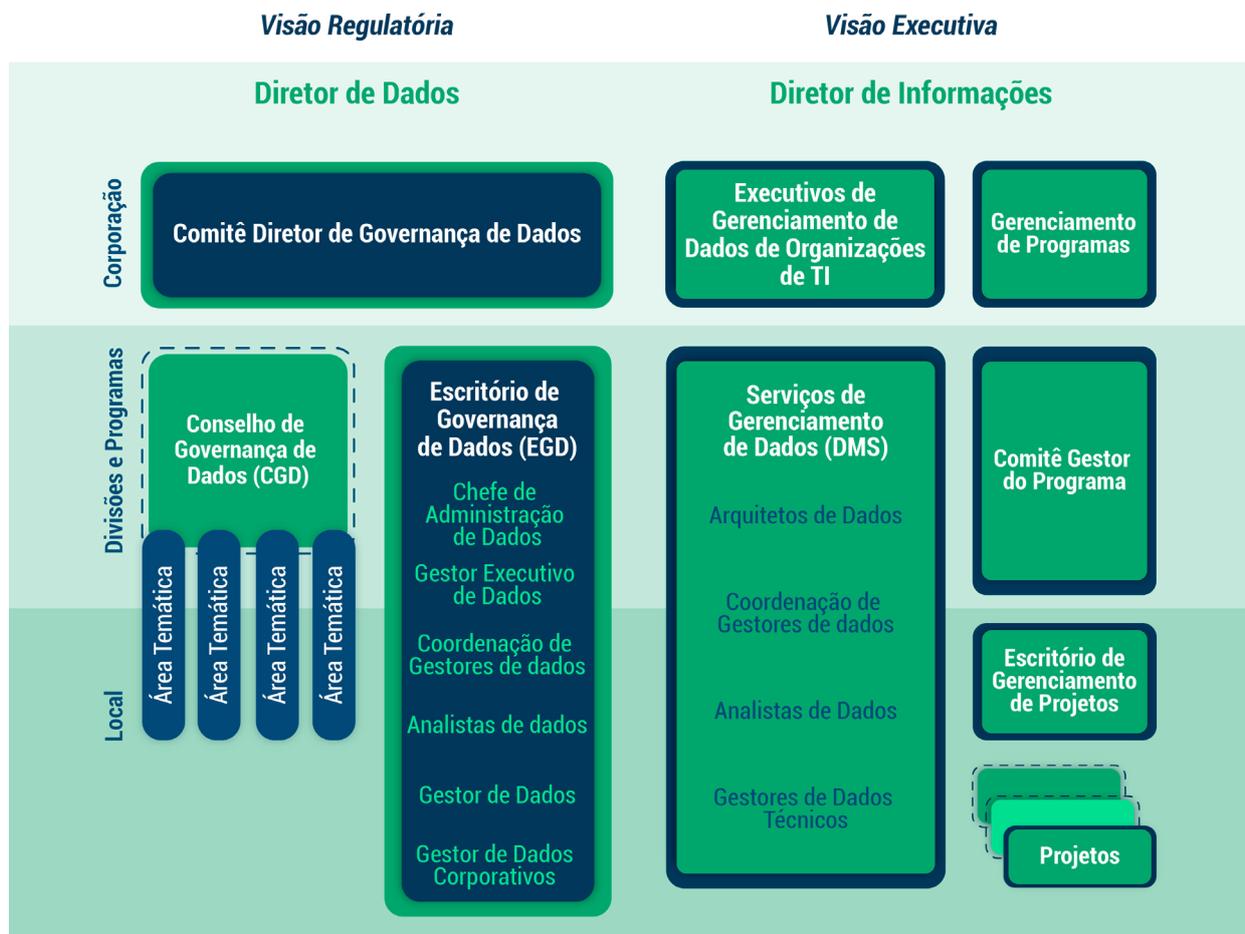


## 1.4 Modelos de Governança de Dados

Cada organização deve adotar um modelo de governança que apoie sua estratégia de negócio. Os modelos diferem em relação à sua estrutura organizacional. Alguns modelos são organizados de forma centralizada, enquanto outros são distribuídos. Todos os modelos precisam de um certo grau de flexibilidade. As organizações também devem estar preparadas para desenvolver seu modelo para enfrentar novos desafios e se adaptar à medida que a cultura organizacional evolui.

A governança de dados pode ser estruturada em camadas e conduzida por comitês, com certo grau de supervisão.

O diagrama a seguir demonstra um modelo genérico de governança de dados, que envolve atividades em vários níveis, bem como a separação das responsabilidades.



Componentes da Governança de Dados em Diferentes Níveis Organizacionais (DMBOK2)

No seguimento da visão regulatória, os gestores de dados são responsáveis por diferentes tipos de dados, dependendo da necessidade da organização. Frequentemente, lideram grupos de trabalho de áreas ou funções que se reportam por meio de um Conselho de Governança de



Dados. No nível estratégico, diversas organizações têm um Comitê Gestor de Governança de Dados. Esse Comitê tem caráter diretivo/deliberativo e monitora a aplicação da política de GD e atua também como Fórum para discussão para dirimir questões sobre gestão de dados e atua na criação de subcomitês para analisar e avaliar soluções.

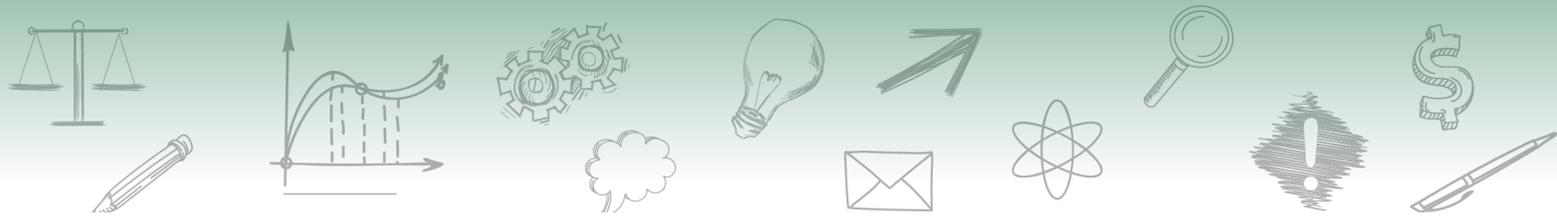
No seguimento da visão executiva, vários programas e vários projetos são implementados por meio de equipes de projeto de gerenciamento de dados.

O próximo diagrama apresenta um tipo de modelo que pode ser implementado conforme as características e restrições de cada organização.



Exemplos de Estrutura Operacional de Governança de Dados Corporativa (DMBOK2)

AT = Área Temática



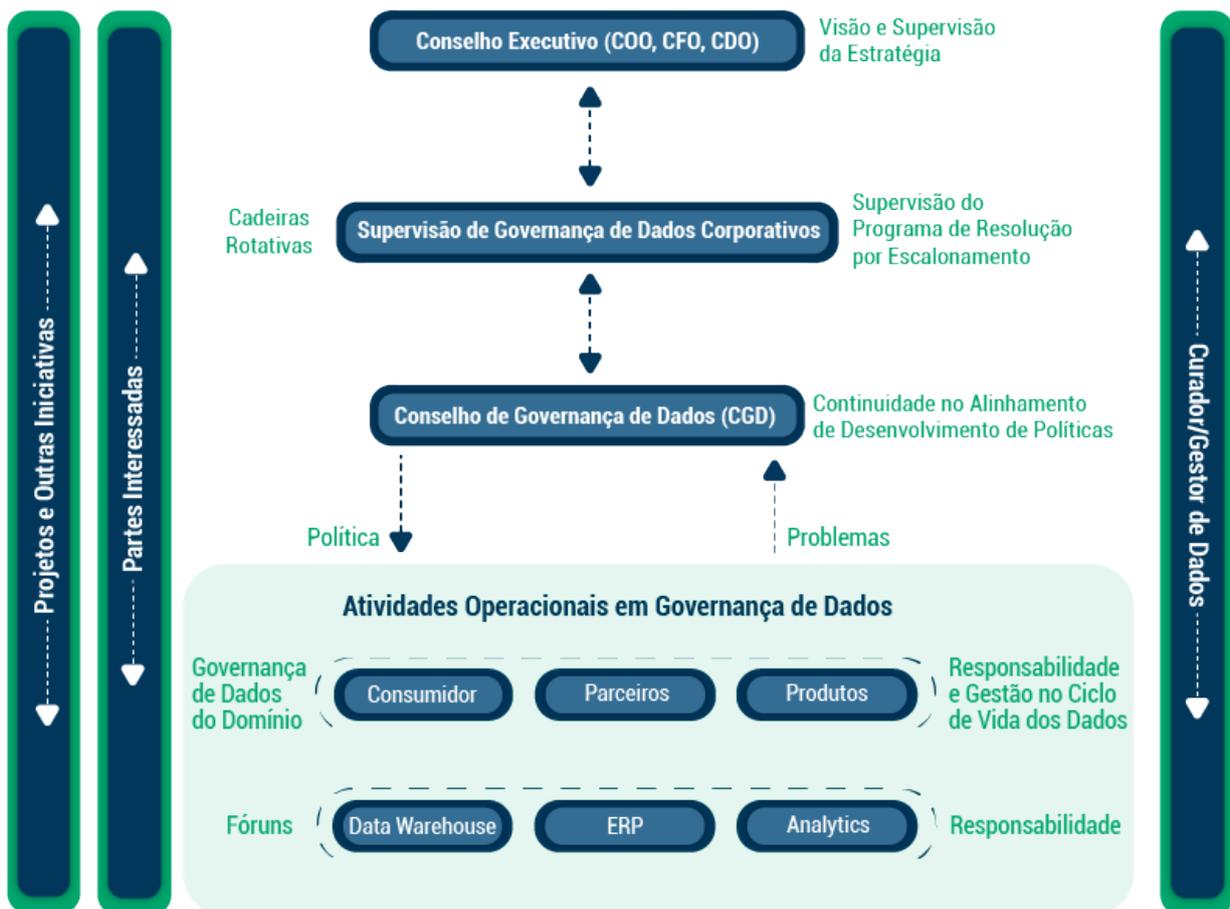
Em um modelo centralizado, a governança de dados supervisiona todas as atividades em todas as áreas de assunto.

Em um modelo replicado, o mesmo modelo operacional e padrões da GD são adotados por cada unidade de negócios.

Em um modelo federado, a governança de dados coordena em parceria com outras unidades de negócio, visando manter as políticas e padrões consistentes.

A governança de dados deve promover um modelo de interação das iniciativas de gerenciamento de dados com atividades de gerenciamento de mudanças.

O diagrama a seguir ilustra um exemplo de como adaptar para atender aos requisitos correspondente à cultura da organização. As funções executivas fornecem supervisão. O DGO (Data Governance Operations) trabalha em domínios. A política é flexível e os problemas são escalonados. Os administradores e as partes interessadas se envolvem em vários níveis.



Exemplo de Modelo Operacional de Governança de Dados (DMBOK2)



## 1.5 Gerenciamento de dados

O *Data Stewards* tem a função de gerenciar ativos de dados em nome de terceiros, conforme interesse da organização.

Os *Data Stewards* representam os interesses das partes interessadas e devem adotar uma perspectiva corporativa e responsável para garantir que os dados corporativos sejam submetidos aos processos de qualidade de dados, possam ser utilizados com eficiência e seguindo as normas de segurança. Os *Data Stewards* são responsáveis pelas atividades de governança de dados.

O foco das atividades de *Data Stewards* está alinhado às políticas e a estratégia de governança de dados de cada organização. As atividades de *Data Stewards* se concentram na seguintes premissas:

- **Criação e gerenciamento de metadados:** padronização, definição e gerenciamento de terminologia de negócios, valores de dados válidos e outros metadados críticos.
- **Regras e padrões de documentação:** definição/documentação de regras de negócios, padrões e regras de qualidade de dados. As expectativas usadas para definir dados de alta qualidade são frequentemente formuladas em termos de regras fundamentadas nos processos de negócios que criam ou consomem dados.
- **Gerenciamento de problemas de qualidade de dados:** Os organizadores geralmente estão envolvidos na identificação, priorização e resolução de problemas relacionados a dados ou na facilitação do processo de resolução.
- **Execução de atividades operacionais de governança de dados:** Responsáveis por garantir que as políticas e iniciativas de governança de dados sejam cumpridas. Influenciam as decisões para garantir que os dados sejam gerenciados de maneira a apoiar os objetivos da organização.

## 1.6 Iniciando o Controle de Dados

A governança de dados permite compartilhar responsabilidades relacionadas a gestão de dados e suas atividades são transversais aos setores da organização. Para isso, é necessário que as partes interessadas internalizem a estratégia e os desafios específicos da gestão dos dados da organização.

As avaliações iniciais permitiram avaliar:

- **A maturidade dos processos de gerenciamento de dados:** como a organização utiliza o potencial humano, processos e tecnologia para gerenciar seus dados. Essa avaliação pode ajudar a determinar os níveis de gestão, os padrões existentes e a identificar oportunidades de melhoria.



- **A capacidade de mudança:** possíveis obstáculos e ações de superação para adoção de ações para gerar comportamentos para implementação de uma governança de dados bem-sucedida.
- **A capacidade colaborativa:** capacidade da organização de colaborar entre funções para administrar e tomar decisões holísticas e consistentes sobre dados.
- **O alinhamento estratégico:** avaliar se a organização possui seus objetivos de gerenciamento de dados alinhados com a estratégia de negócio.
- **A qualidade dos dados:** identificar se os dados críticos e pontos problemáticos sobre riscos associados aos dados e aos processos de negócio.
- **A conformidade regulatória:** a relação dos riscos com os requisitos de conformidade e como são gerenciados. Identificar controles e monitoramento que possam melhorar a capacidade da organização de cumprir com os regulamentos.

As avaliações de maturidade, qualidade dos dados e conformidade devem identificar pontos de melhoria e orientar uma estratégia que defina o escopo e a abordagem dos esforços de governança de dados em relação aos objetivos de negócio. A estratégia deve ser definida por meio de:

- Um instrumento formal que estabeleça metas e princípios.
- Uma matriz de responsabilidades.
- Um plano de implementação estratégico e operacional que descreva:
- O objetivo das atividades de governança de dados e como elas serão incorporadas nos processos de TI.
- O conjunto de iniciativas para melhorar os recursos de gerenciamento e da qualidade dos dados.
- Os benefícios esperados.
- A definição dos indicadores que demonstre os resultados esperados.

Depois que a estratégia é definida, segue-se:

- A definição das políticas.
- A subscrição dos projetos de melhoria de dados.
- O gerenciamento de mudanças para impulsionar a adoção dos comportamentos desejáveis.



- O gerenciamento de problemas e conflitos que podem surgir durante o processo de implementação.

## 1.7 Governança de Dados Sustentável

A Governança de Dados orienta o gerenciamento de dados, estabelecendo políticas e práticas para sua implementação. Como essas práticas devem ser realizadas por outras áreas, os princípios devem ser incorporados ao ciclo de vida de gerenciamento de dados e às atividades fundamentais.

Um programa de governança de dados requer:

- Estabelecer uma estratégia que se alinhe e suporte a estratégia de negócio.
- Definir e aplicar políticas para mudança de cultura organizacional.
- Definir padrões para processos de qualidade dos dados.
- Fornecer gerenciamento de dados críticos.
- Garantir que a organização esteja em conformidade com as normas e regulamentos relacionados a tratamento de dados.
- Gerenciar riscos relacionados.

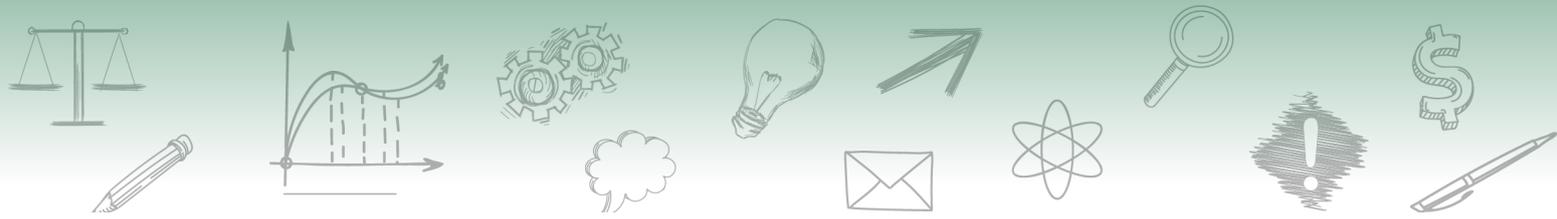
Um programa de governança de dados implicará na mudança da cultura, priorizando ações que incidem na curva de maturidade:

- Patrocinar projetos/ações de gerenciamento de dados.
- Padronizando a avaliação de ativos de dados.
- Estabelecer planos de comunicação contínuos sobre a implementação dos processos de gerenciamento de dados.

## 2. Planejamento e design no gerenciamento do ciclo de vida dos dados

As atividades de gerenciamento do ciclo de vida dos dados se concentram no planejamento e design de dados, permitindo o uso, a manutenção para atender às metas organizacionais. Arquitetos de dados e modeladores de dados são os que planejam e projetam dados.

Este tópico descreverá:



- O papel da arquitetura corporativa no planejamento e design da organização
- A função crítica da arquitetura de dados no gerenciamento de dados
- Os objetivos e artefatos associados à modelagem de dados

## 2.1 Arquitetura de Negócio

Arquitetura refere-se a um arranjo organizado de elementos destinados a otimizar a função, o desempenho, a viabilidade, o custo e a estética de uma estrutura ou sistema geral. O termo arquitetura foi adotado para descrever várias facetas do design de sistemas de informação. Mesmo em pequenas organizações, a tecnologia da informação é complicada.

Artefatos arquitetônicos e documentação que descrevem sistemas e fluxos de dados mostram às pessoas como sistemas, processos e dados funcionam juntos. Uma abordagem estratégica da arquitetura permite que uma organização tome melhores decisões sobre seus sistemas e dados.

Uma prática de arquitetura corporativa bem gerenciada pode ajudar uma organização a entender o estado atual de seus sistemas, promover mudanças desejáveis em relação ao estado futuro, habilitar a conformidade regulatória e melhorar a eficácia. O gerenciamento eficaz de dados e os sistemas em que os dados são armazenados e usados é um objetivo comum da amplitude das disciplinas de arquitetura.

A prática de arquitetura é realizada em diferentes níveis dentro de uma organização (incluindo organização, domínio ou projeto) e com diferentes áreas de foco (por exemplo, infraestrutura, aplicações ou dados).

A tabela abaixo descreve e compara os domínios de arquitetura. Arquitetos de diferentes domínios devem atender aos requisitos de desenvolvimento de forma colaborativa, pois cada domínio influencia os outros domínios.

Tipo de Arquitetura	Objetivo	Elementos	Dependências	Papéis
<b>Arquitetura de Negócios Corporativos</b>	Identificar como uma corporação cria valor para clientes/usuários e outras partes interessadas.	Modelos de negócio, processos, serviços de recursos, eventos, estratégias, vocabulário.	Estabelece requisitos para os outros domínios.	Arquitetos e analistas de negócio, Gestores de dados de negócios.



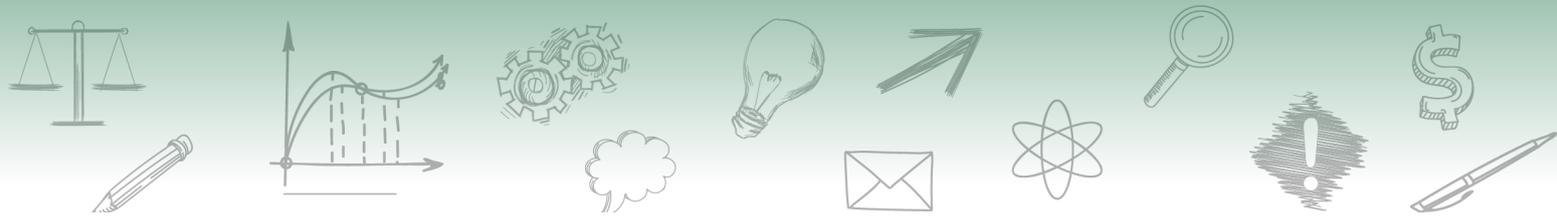
<b>Arquitetura de Dados Corporativos</b>	Descrever como os dados devem ser organizados e gerenciados.	Modelos de dados, definições de dados, especificações de mapeamento de dados, fluxos de dados, APIs de dados estruturados.	Gerencia dados criados e exigidos pela arquitetura de negócios.	Arquitetos e modeladores de dados, Gestores de dados.
<b>Arquitetura de Aplicações Corporativas</b>	Descrever a estrutura e a funcionalidade das aplicações em uma organização.	Sistemas de negócios, pacotes de software, bancos de dados.	Atua nos dados especificados de acordo com os requisitos de negócios.	Arquitetos de aplicações.
<b>Arquitetura de Tecnologia Corporativa</b>	Descrever a tecnologia física necessária para permitir que os sistemas funcionem e agreguem valor.	Plataformas técnicas, redes, ferramentas de integração de segurança.	Hospeda e executa a arquitetura de aplicações.	Arquitetos de infraestrutura.

## 2.2 O Quadro Zachman

Uma estrutura de arquitetura é uma estrutura fundamental usada para desenvolver uma ampla variedade de arquiteturas relacionadas. Ela fornece uma maneira de pensar e entender a arquitetura e representa uma 'arquitetura para arquitetura' geral. As estruturas descritas na forma do diagrama permitem que profissionais sem a formação específica compreendam os relacionamentos (se não as diferenças detalhadas) entre esses conceitos.

A estrutura de arquitetura corporativa mais conhecida, o Zachman Framework, foi desenvolvida por John A. Zachman na década de 1980. Continuou a evoluir. Zachman reconheceu que, ao criar edifícios, aviões, empresas, cadeias de valor, projetos ou sistemas, existem muitas partes interessadas e cada uma tem uma perspectiva diferente sobre arquitetura. Ele aplicou esse conceito aos requisitos para diferentes tipos e níveis de arquitetura em uma organização.

O Zachman Framework é representado por uma matriz 6x6 que resume o conjunto completo de modelos necessários para descrever uma organização e os relacionamentos entre eles. O Framework não define como criar os modelos, mas demonstra quais modelos devem existir.



	O que?	Como?	Onde?	Quem?	Quando?	Por quê?	
Executivo	Identificação do Inventário	Processo de Identificação	Identificação e Distribuição	Identificação e Responsabilidade	Identificação do Tempo	Identificação e Motivação	Contexto do Escopo
Gestão de Negócios	Identificação do Inventário	Definição de Processo	Definição e Distribuição	Definição e Responsabilidade	Definição do Tempo	Definição e Motivação	Conceitos de Negócios
Arquiteto	Inventário de Representação	Representação do Processo	Representação e Distribuição	Responsabilidade da Representação	Representação do Tempo	Representação e Motivação	Lógica do Sistema
Engenheiro	Especificação do Inventário	Especificação do Processo	Especificação e Distribuição	Especificação e Responsabilidade	Especificação do Tempo	Especificação e Motivação	Tecnologia Física
Técnico	Configuração do Inventário	Configuração do Processo	Configuração e Distribuição	Configuração e Responsabilidade	Configuração do período	Configuração e Motivação	Ferramentas e Componentes
Corporação	Instancias de Inventário	Instancias de Processo	Instancias de Distribuição	Instancias de Responsabilidade	Instancias de Tempo	Instancias e Motivação	Instancias Operacionais
	Conjunto de Inventários	Fluxo de Processos	Distribuição de Redes	Atribuições de Responsabilidade	Ciclo de Tempo	Intenções e Motivação	

Framework simplificado de Zachman (DMBOK2)

Atenção à estrutura de Zachman! Ela resume as respostas a um conjunto simples de perguntas (ou seja, o que, como, onde, quem, quando, por que) que podem ser feitas pelas partes interessadas com diferentes perspectivas. São as seguintes perspectivas:

- **executiva (contexto de negócios):** Lista os elementos de negócios que definem o escopo nos modelos de identificação.
- **de gerenciamento de negócios (conceitos de negócios):** Esclarecimento das relações entre os conceitos de negócios definidos pelos líderes executivos como proprietários nos modelos de definição.
- **do arquiteto (lógica de negócios):** Modelos lógicos do sistema detalhando os requisitos do sistema e o design sem restrições representados pelos arquitetos como designers nos modelos de representação.
- **do engenheiro (física dos negócios):** Modelos físicos otimizando o design para implementação para uso específico sob as restrições de tecnologia, pessoas, custos e prazos específicos especificados pelos engenheiros como construtores nos modelos de especificação.



- **do técnico (montagens de componentes):** Uma visão fora de contexto, específica da tecnologia, de como os componentes são montados e operados, configurados pelos técnicos como implementadores nos modelos de configuração.
- **do usuário (classes de operações):** Instâncias de funcionamento reais usadas por Trabalhadores como Participantes. Não há modelos nessa perspectiva.

A estrutura então identifica quais tipos de artefatos de arquitetura são necessários para responder a essas perguntas fundamentais.

## 2.3 Arquitetura de Dados

A arquitetura de dados é fundamental para o gerenciamento de dados. Como a maioria das organizações possui mais dados do que as pessoas podem compreender, é necessário representar dados organizacionais em diferentes níveis de abstração para que a alta Administração possa analisá-los e tomar decisões sobre eles.

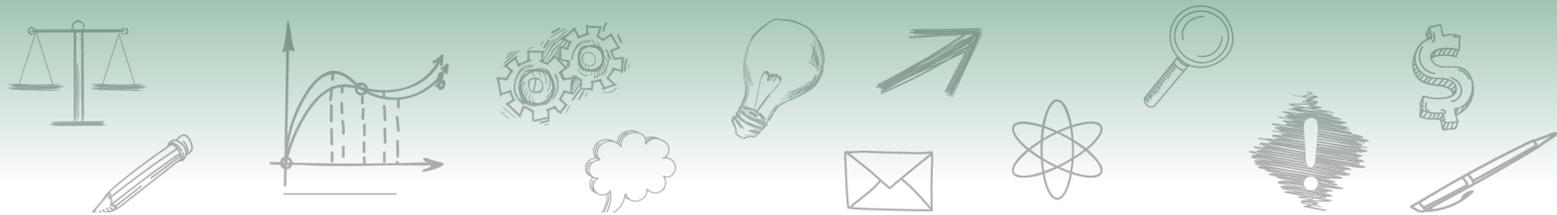
A disciplina especializada da arquitetura de dados pode ser entendida sob alguns pontos de vista, apresentados a seguir:

- **Resultados da arquitetura de dados**, tais como modelos, definições e fluxos de dados em vários níveis (geralmente referidos como artefatos da arquitetura de dados).
- **Atividades de arquitetura de dados** para formar, implantar e cumprir intenções de arquitetura de dados.
- **Comportamento da arquitetura de dados**, assim como colaboração e habilidades entre as várias funções que afetam a arquitetura de dados da organização.

A arquitetura de dados de uma organização é descrita por uma coleção integrada de documentos de projeto mestre em diferentes níveis de abstração, incluindo padrões que governam como os dados são coletados, armazenados, organizados, usados e removidos. Também é classificada por descrições de todos os contêineres e caminhos que os dados percorrem nos sistemas de uma organização.

Os artefatos da arquitetura de dados incluem especificações usadas para descrever o estado existente, definir requisitos de dados, orientar a integração de dados e controlar ativos de dados, conforme apresentado em uma estratégia de dados.

O documento de design da arquitetura de dados mais detalhado é um modelo formal de dados corporativos, contendo nomes de dados, definições abrangentes de dados e metadados, entidades e relacionamentos conceituais e lógicos e regras de negócios. Modelos de dados físicos estão incluídos, mas como um produto de modelagem e design de dados ao invés de uma arquitetura de dados.



A arquitetura de dados agrega mais valor quando suporta as necessidades de gestão de dados de toda a organização. A arquitetura de dados corporativos define termos e designs padrão para os elementos importantes para toda a organização.

O design de uma arquitetura de dados corporativos inclui a representação dos dados de negócio como tais, englobando a coleta, armazenamento, integração, movimentação e distribuição de dados. A arquitetura de dados corporativos permite padronização e integração consistentes de dados em toda a organização.

A arquitetura de dados servirá como uma ponte entre a estratégia de negócios e a execução da tecnologia. Vale alertar que como parte da arquitetura corporativa os arquitetos de dados devem:

- Preparar estrategicamente as organizações para evoluir rapidamente seus produtos, serviços e dados para aproveitar as oportunidades de negócios inerentes às tecnologias emergentes.
- Traduzir necessidades de negócios em requisitos de dados e de sistema, para que os processos tenham consistentemente os dados necessários.
- Gerenciar entrega complexa de dados e informações em toda a organização.
- Facilitar o alinhamento entre negócios e TI.
- Agir como agentes de mudança e transformação.

Esses direcionadores de negócios devem influenciar as medidas do valor da arquitetura de dados.

## 2.4 Artefatos de arquitetura de dados

À medida que os dados fluem dentro de uma organização, por meio de feeds ou interfaces, eles são protegidos, integrados, armazenados, registrados, catalogados, compartilhados, relatados, analisados e entregues às partes interessadas.

Ao longo do caminho, os dados podem ser verificados, aprimorados, vinculados, certificados, agregados, anonimizados e usados para análises até serem arquivados ou eliminados. As descrições da arquitetura de dados corporativos devem, portanto, incluir modelos de dados corporativos (por exemplo, estruturas e especificações de dados), bem como projetos de fluxo de dados.

Os arquitetos de dados criam e mantêm o conhecimento organizacional sobre os dados e os sistemas pelos quais eles se movem. Esse conhecimento permite que uma organização gerencie seus dados como um ativo e aumente o valor que obtém de seus dados, identificando oportunidades para uso de dados, redução de custos e mitigação de riscos.



Os arquitetos buscam projetar de maneira a agregar valor à organização. Esse valor é obtido por meio de presença técnica ideal, eficiências operacionais e de projetos e maior capacidade da organização de usar seus dados. Para chegar lá, são necessários: bom design, planejamento e capacidade de garantir que os projetos e planos sejam executados de maneira eficaz.

Nesse sentido, serão apresentados, a seguir, alguns artefatos da arquitetura de dados.

## **Modelo de Dados Corporativos (Enterprise Data Model - EDM)**

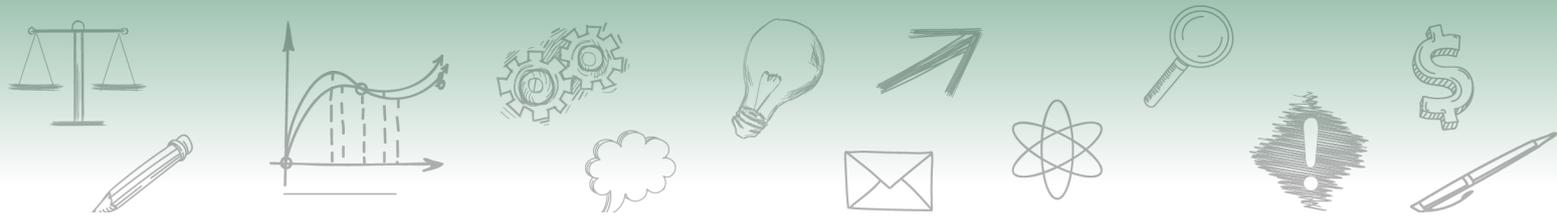
O Enterprise Data Model - EDM é um modelo de dados conceitual ou lógico holístico, que independente da implementação, fornece uma visão comum e consistente dos dados da organização. Um EDM inclui as principais entidades de dados corporativos (ou seja, conceitos de negócios), seus relacionamentos, regras críticas de negócios orientadoras e alguns atributos críticos. Estabelece a base para todos os dados e projetos relacionados a dados.

Qualquer modelo de dados no nível de projeto deve ser baseado no EDM. O EDM deve ser revisado pelas partes interessadas em concordância com a estratégia de negócio da organização.

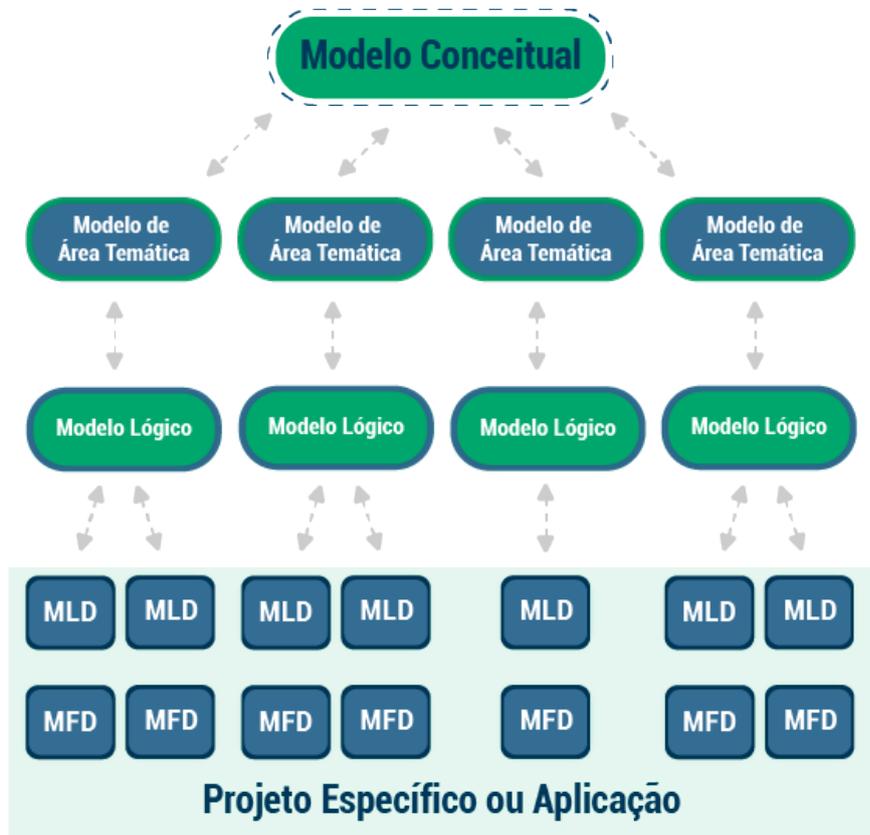
Uma organização que reconhece a necessidade de um modelo de dados corporativo deve decidir quanto tempo e esforço pode dedicar à sua construção.

Os EDMs podem ser criados em diferentes níveis de detalhes, portanto, a disponibilidade de recursos influenciará o escopo inicial. Com o tempo, conforme as necessidades da organização exigem, o escopo e o nível de detalhes capturados em um modelo de dados corporativos geralmente se expandem. Os modelos de dados corporativos mais bem-sucedidos são criados de forma incremental e iterativa, usando camadas.

A imagem a seguir apresenta os diferentes tipos de modelos e mostra como os modelos conceituais são vinculados aos modelos de dados de aplicações lógicas e físicas.



## Modelo de Dados Corporativo



**Diagrama I:**  
12-20 áreas temáticas de negócios significativas com relacionamentos

**Diagrama I +**  
por área temática: mais de 50 entidades significativas dentro da área temática com relacionamentos

**Para cada modelo lógico da área temática:** aumente os detalhes adicionando atributos e entidades e relacionamentos menos significativo

Limitar o escopo ao tema e aos relacionamentos externos da etapa I

Limitar o escopo aos modelos físicos e áreas temáticas do relacionamentos

MLD = Modelo Lógico de Dados  
MFD = Modelo Físico de Dados

Modelo de Dados Corporativos (DMBOK2)

Com relação à imagem acima, que apresenta o Modelo de Dados Corporativos, vale ressaltar que ela permite reconhecer:

- a perspectiva conceitual sobre as áreas temáticas da organização.
- as entidades e relacionamentos para cada área temática.
- perspectiva lógica detalhada e parcialmente atribuída a cada uma dessas áreas temáticas.
- os modelos lógicos e físicos específicos de uma aplicação ou projeto.

Todos os níveis fazem parte desse modelo e as vinculações criam caminhos para rastrear uma entidade de cima para baixo e entre modelos no mesmo nível.



## Design de Fluxo de Dados

O design do fluxo de dados define os requisitos e o plano mestre para armazenamento e processamento em bancos de dados, aplicações, plataformas e redes (os componentes). Esses fluxos de dados mapeiam a movimentação de dados para processos, locais, funções de negócios e componentes técnicos.

Os fluxos de dados são um tipo de documentação de linhagem de dados que descreve como os dados transitam pelos processos e sistemas de negócios.

Os fluxos de dados de ponta a ponta ilustram onde os dados se originaram, onde são armazenados e usados e como são transformados à medida que se movem dentro e entre diversos processos e sistemas.

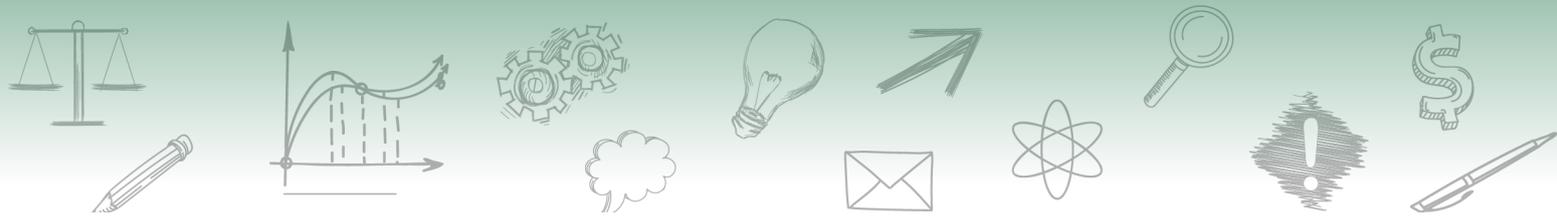
A análise de linhagem de dados pode ajudar a explicar o estado dos dados em um determinado ponto no fluxo.

Ressalta-se que os fluxos de dados mapeiam e documentam os relacionamentos entre:

- Aplicações em um processo de negócios.
- Armazenamentos de dados ou bancos de dados em um ambiente.
- Segmentos de rede (úteis para mapeamento de segurança).
- Funções de negócio, representando quais funções são responsáveis por criar, atualizar, usar e excluir dados (CRUD – Acrônimo de *Create*, *Read*, *Update* e *Delete*, na língua inglesa).
- Locais onde ocorrem as diferenças.

Os fluxos de dados podem ser documentados em diferentes níveis de detalhe: área de assunto, entidade ou mesmo no nível do atributo. Os sistemas podem ser representados por segmentos de rede, plataformas, conjuntos de aplicações comuns ou servidores individuais.

Os fluxos de dados podem ser representados por matrizes bidimensionais e diagrama de fluxo de dados:



- **Matrizes Bidimensionais**

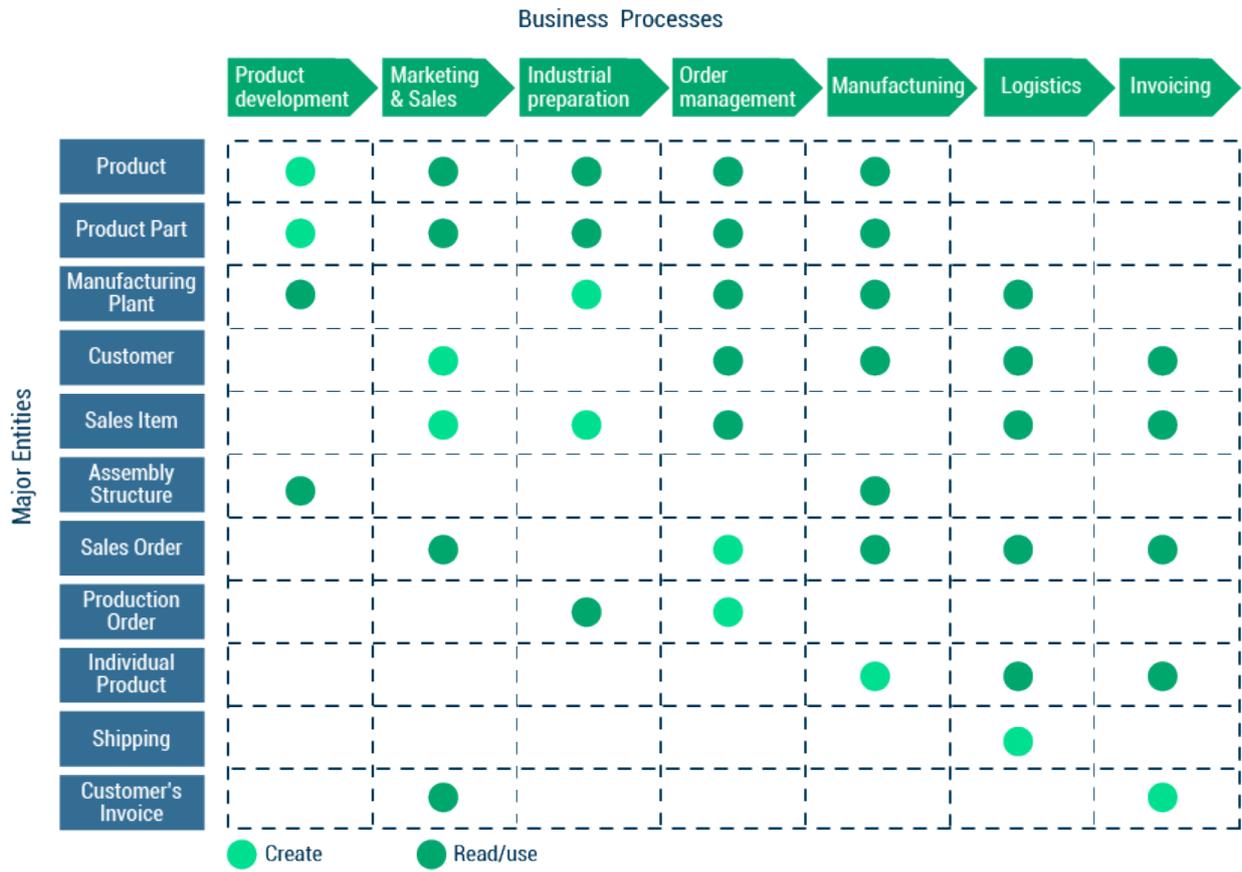


Figura 15: Fluxo de dados representado em uma Matriz Bidimensional (DMBOK2)



- Diagramas de Fluxo de Dados (DFD)

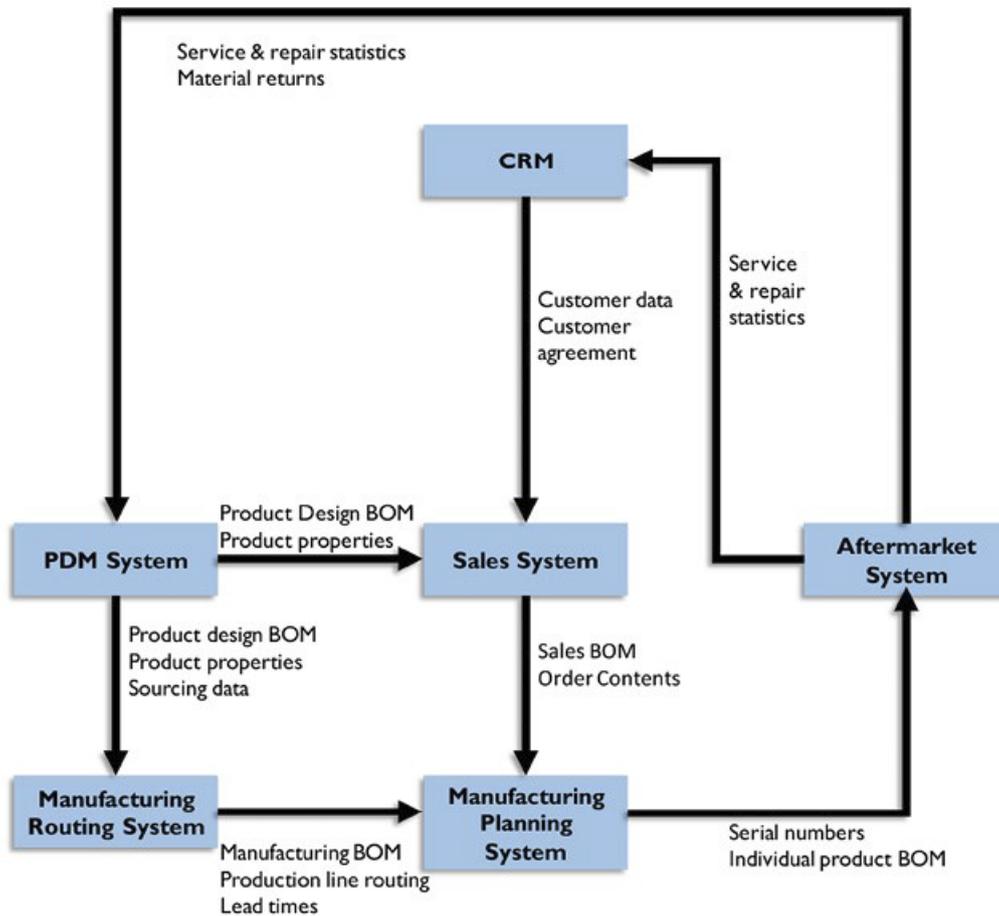


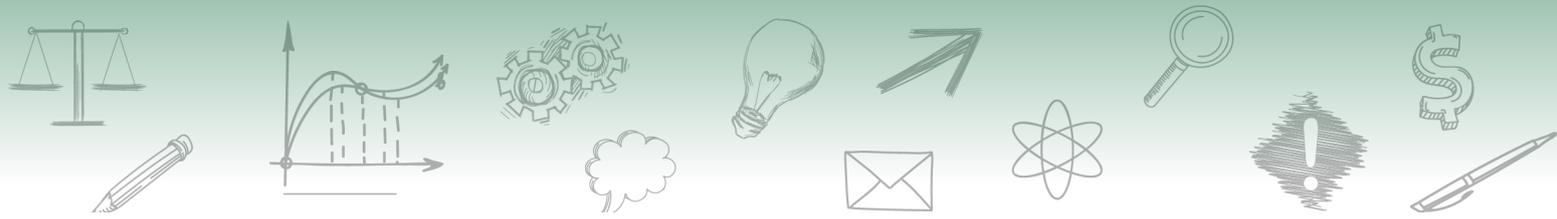
Figura 16: Exemplo de Diagrama de Fluxo de Dados (DMBOK2)

## DESTAQUE

O Modelo de Dados Corporativos e o Design do Fluxo de Dados precisam estar bem alinhados. Como mencionado, ambos precisam refletir o estado atual em relação à perspectiva da arquitetura de dados e também o estado de transição em relação à perspectiva do projeto.

### Arquitetura, Qualidade e Inovação de Gerenciamento de Dados

Os dados e a arquitetura corporativa lidam com a complexidade de dois indicadores, que pedem abordagens distintas, conforme apresentado na tabela a seguir.



Direcionadores/Drivers	Foco	Abordagem
Orientação para a qualidade	<p>Concentre-se em melhorar a execução nos ciclos de negócios e desenvolvimento de TI. A menos que a arquitetura seja gerenciada, a arquitetura se deteriorará.</p> <p>Os sistemas gradualmente se tornarão mais complexos e inflexíveis, criando riscos para uma organização. Entrega descontrolada de dados, cópias de dados e relacionamentos de "espaguete" de interface tornam as organizações menos eficientes e reduzem a confiança nos dados.</p>	<p>Alinha-se ao trabalho tradicional de arquitetura de dados, no qual as melhorias na qualidade da arquitetura são realizadas de forma incremental por meio da conexão do arquiteto com os projetos.</p> <p>Normalmente, o arquiteto mantém em mente toda a arquitetura e se concentra nos objetivos de longo prazo diretamente conectados à governança, padronização e desenvolvimento estruturado.</p>
Orientação para a inovação	<p>Concentre-se na transformação dos negócios e da TI para atender novas expectativas e oportunidades.</p> <p>Impulsionar a inovação com tecnologias disruptivas e uso de dados tornou-se um papel do arquiteto corporativo moderno.</p>	<p>Pode ter uma perspectiva de curto prazo e usar lógica de negócios não comprovada e tecnologias de ponta.</p> <p>Essa orientação geralmente exige que os arquitetos entrem em contato com pessoas da organização com as quais os profissionais de TI geralmente não interagem; por exemplo, representantes de desenvolvimento de produtos e designers de negócios.</p>

Ao se trabalhar na arquitetura corporativa ou como uma equipe de arquitetura de dados, os arquitetos de dados são responsáveis por desenvolver um roteiro, gerenciar requisitos de dados corporativos nos projetos e integrar-se à arquitetura corporativa geral.

O sucesso depende da definição e adesão aos padrões e da criação e manutenção de artefatos de arquitetura úteis e utilizáveis. Uma prática de arquitetura disciplinada pode melhorar a eficiência e a qualidade, criando soluções reutilizáveis e extensíveis.



## 2.5 Modelagem de dados

Um *modelo* é uma representação de algo que existe ou um padrão para algo a ser feito. Mapas, organogramas e plantas de construção são exemplos de modelos em uso todos os dias. Os diagramas de modelos usam símbolos padrão que permitem entender seu conteúdo.

A modelagem de dados é o processo de descoberta, análise e escopo de requisitos de dados e, em seguida, representação e comunicação desses requisitos de dados de uma forma precisa chamada modelo de dados. A modelagem de dados é um componente crítico do gerenciamento de dados. O processo de modelagem exige que as organizações descubram e documentem como seus dados se encaixam. Modelos de dados permitem que uma organização entenda seus ativos de dados.

Os modelos de dados incluem e contêm metadados essenciais para os consumidores de dados. Boa parte desses metadados descobertos durante o processo de modelagem é essencial para outras funções de gerenciamento de dados, como, por exemplo, definições para controle de dados e linhagem para armazenamento e análise de dados.

Um modelo de dados descreve os dados de uma organização como a organização os entende ou como a organização deseja que sejam. Um modelo de dados contém um conjunto de símbolos com rótulos de texto que tentam representar visualmente os requisitos de dados comunicados ao modelador de dados, para um conjunto específico de dados que pode variar em tamanho, de pequeno (para um projeto) a grande (para uma organização).

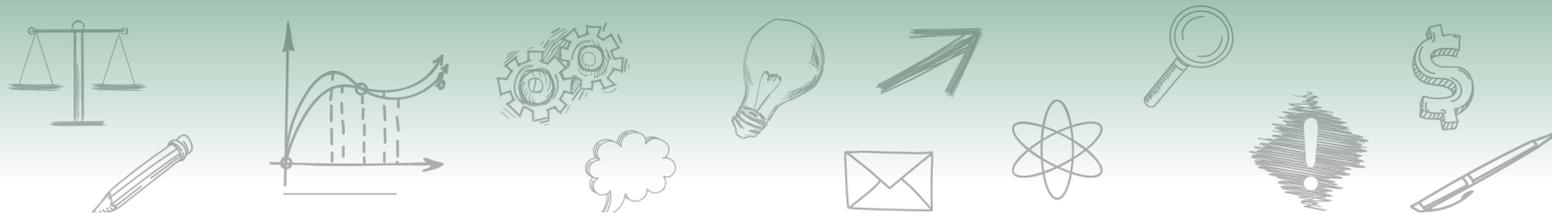
O modelo é, portanto, uma forma de documentação para requisitos e definições de dados resultantes do processo de modelagem. Os modelos de dados são o principal meio usado para comunicar os requisitos de dados da organização para a TI e, dentro da TI, de analistas e arquitetos, para designers e desenvolvedores de bancos de dados.

Atenção, os modelos de dados são críticos para o gerenciamento eficaz dos dados, pois:

- Fornecem um vocabulário comum sobre dados.
- Capturam e documentam o conhecimento explícito (metadados) sobre os dados e sistemas de uma organização.
- Servem como principal ferramenta de comunicação durante os projetos.
- Fornecem o ponto de partida para personalização, integração ou mesmo substituição de uma aplicação.

### Objetivos de modelagem de dados

O objetivo da modelagem de dados é confirmar e documentar o entendimento de diferentes perspectivas sobre os dados. Esse entendimento leva a aplicações e dados que se alinham mais



com os requisitos de negócios atuais e futuros. Esse entendimento também cria uma base para concluir com êxito iniciativas de escopo amplo, como Master Data Management e programas de governança de dados.

A modelagem de dados adequada leva a custos de suporte mais baixos e aumenta as oportunidades de reutilização para iniciativas futuras, reduzindo assim os custos de construção de novas aplicações. Além disso, os próprios modelos de dados são uma forma importante de metadados.

Alertamos que, confirmar e documentar a compreensão de diferentes perspectivas facilita à:

- **Formalização** - Um modelo de dados documenta uma definição concisa de estruturas e relacionamentos de dados. Ele permite avaliar como os dados são afetados pelas regras de negócios implementadas, para os estados atuais (como estão) ou os estados de destino desejados. A definição formal impõe uma estrutura disciplinada aos dados que reduz a possibilidade de anomalias nos dados ao acessar e persistir dados. Ao ilustrar as estruturas e os relacionamentos nos dados, um modelo de dados facilita o consumo dos dados.
- **Definição do escopo** - Um modelo de dados pode ajudar a explicar os limites do contexto de dados e implementação de pacotes de aplicativos, projetos, iniciativas ou sistemas existentes.
- **Retenção/documentação de conhecimento** - Um modelo de dados pode preservar a memória corporativa em relação a um sistema ou projeto capturando conhecimento de forma explícita. Ele serve como documentação para futuros projetos a serem usados como estão.

Os modelos de dados nos ajudam a entender uma organização ou área de negócios, um aplicativo existente ou o impacto da modificação de uma estrutura de dados existente. O modelo de dados se torna um mapa reutilizável para ajudar profissionais de negócios, gerentes de projeto, analistas, modeladores e desenvolvedores a entender a estrutura de dados no ambiente. Da mesma maneira que o cartógrafo aprendeu e documentou um cenário geográfico para outros usarem na navegação, o modelador permite que outros entendam um cenário de informações.

## **Blocos de construção**

Existem diferentes variedades de modelos de dados, incluindo relacionais, dimensionais etc. Os profissionais de design utilizam tipos apropriados de modelos com base nas necessidades da organização, nos dados que estão sendo modelados e no sistema para o qual o modelo está sendo desenvolvido. Cada espécie de modelo utiliza convenções visuais diferentes para capturar informações.

Os modelos também diferem com base no nível de abstração das informações que representam (conceitual com um alto nível de abstração; lógico com um nível médio de abstração; e físico que



representa um sistema específico ou instanciação de dados). Entretanto, todos os modelos usam os mesmos blocos de construção: entidades, relacionamentos, atributos e domínios.

Como líder em sua organização, não é necessário que você possa ler modelos de dados. No entanto, é útil você entender como eles descrevem dados. As definições aqui fornecem um panorama de como os modelos de dados funcionam.

## Entidade

Fora da modelagem de dados, a definição de entidade é algo que existe separado de outras coisas. Na modelagem de dados, uma entidade é algo sobre o qual uma organização coleta informações. Às vezes, as entidades são chamadas de 'substantivos de uma organização. Nos modelos de dados relacionais, as entidades são as caixas que identificam o conceito que está sendo modelado.

Uma entidade pode ser pensada como a resposta a esses questionamentos: Quem? O que? Quando? Onde? Por quê? Como? Pode também ser definida por uma combinação dessas perguntas.

## Relação

Um relacionamento é uma associação entre entidades. Ele captura as interações de alto nível entre entidades conceituais, as interações detalhadas entre entidades lógicas e as restrições entre entidades físicas. Os relacionamentos são mostrados como linhas no diagrama de modelagem de dados.

Em uma relação entre duas entidades, a cardinalidade captura quantas de uma entidade (instâncias de entidade) participam do relacionamento com quantas da outra entidade. Por exemplo, uma organização pode ter um ou vários funcionários.

## Cardinalidade

A cardinalidade é representada pelos símbolos que aparecem nas duas extremidades de uma linha de relacionamento. Para cardinalidade, as opções são simples: zero, um ou muitos ("muitos" referentes a qualquer número maior que um). Cada lado de um relacionamento pode ter qualquer combinação de zero, um ou muitos.

O diagrama abaixo apresenta diferentes relações de cardinalidade. Uma organização emprega um ou mais funcionários. Um funcionário pode suportar zero, um ou muitos dependentes. Mas um Funcionário tem um e apenas um Trabalho durante um período de tempo.

Os relacionamentos de cardinalidade são uma maneira de capturar regras e expectativas relacionadas aos dados. Se os dados mostrarem que um Funcionário mantém mais de um Trabalho durante um período de tempo definido, há um erro nos dados ou a organização está violando uma regra.

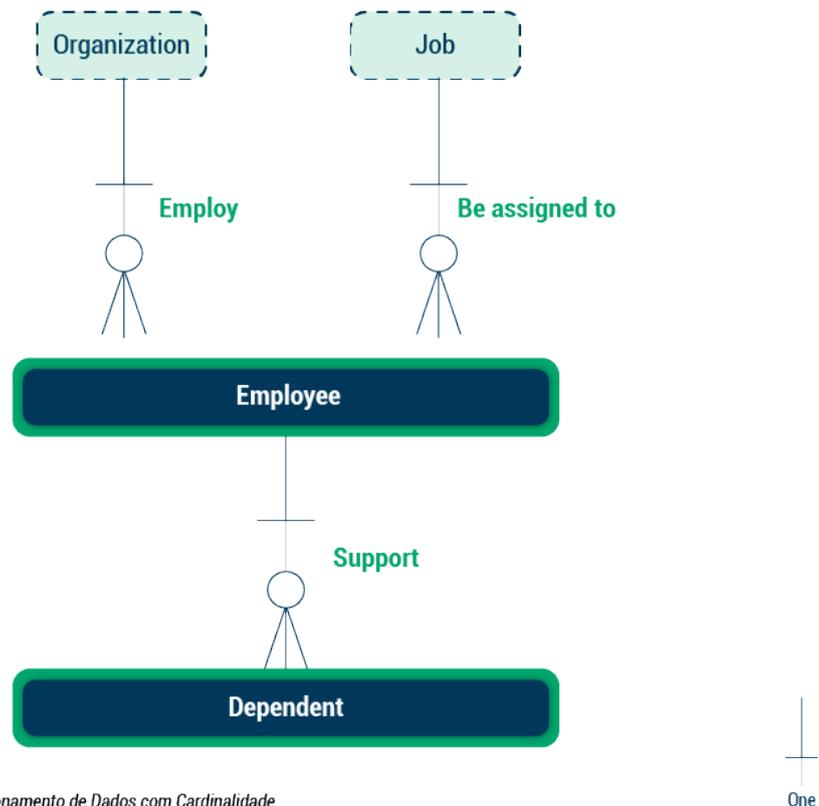
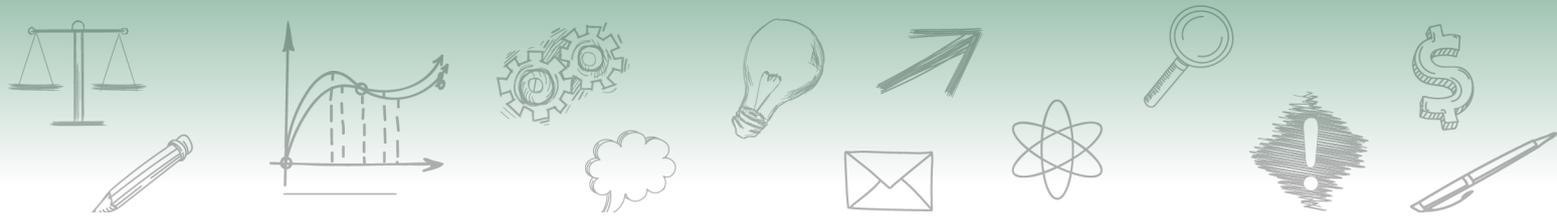


Figura 17: Modelo de Relacionamento de Dados com Cardinalidade

## Atributo

Um atributo é uma propriedade que identifica, descreve ou mede uma entidade. O correspondente físico de um atributo em uma entidade é uma coluna, campo, tag ou nó em uma tabela, visualização, documento, gráfico ou arquivo.

No exemplo abaixo, a entidade *Organization* possui os atributos de ID fiscal da organização, Número de telefone da organização e Nome da organização. O funcionário possui os atributos: Número do funcionário, Nome do funcionário, Sobrenome do funcionário e Data de nascimento do funcionário. Os detalhes Dependente e Trabalho têm atributos que descrevem suas características.

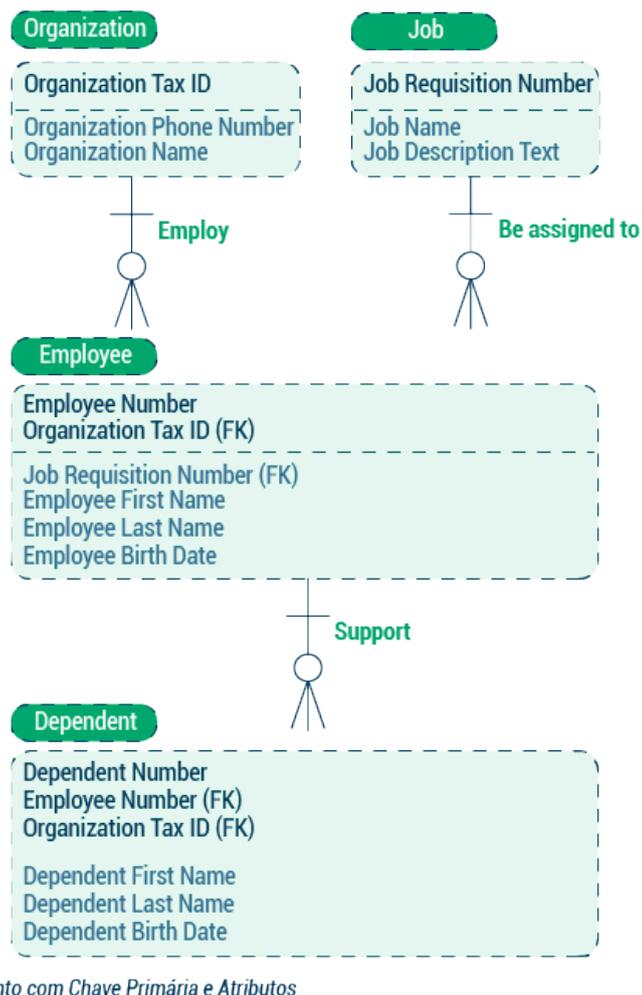


Figura 18: Modelo de Relacionamento com Chave Primária e Atributos

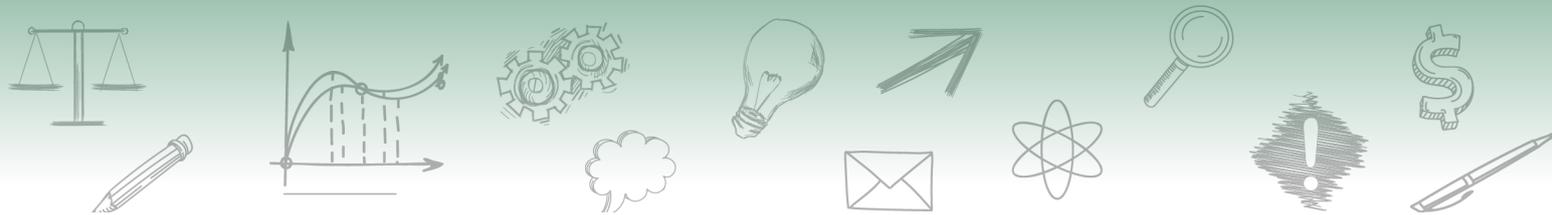


## Domínio

Na modelagem de dados, um domínio é o conjunto completo de valores possíveis aos quais um atributo pode ser atribuído. Um domínio fornece um meio de padronizar as características dos atributos e restringe os dados que podem ser preenchidos no campo.

Por exemplo, o domínio Data, que contém todas as datas válidas possíveis, pode ser atribuído a qualquer atributo de data em um modelo de dados lógico ou a colunas/campos de data em um modelo de dados físico, como:

- EmployeeHireDate
- OrderEntryDate
- ClaimSubmitDate
- CourseStartDate



Os domínios são críticos para entender a qualidade dos dados. Todos os valores dentro do domínio são válidos. Aqueles fora do domínio são chamados de valores inválidos. Um atributo não deve conter valores fora do domínio atribuído. O domínio para *EmployeeHireDate* pode ser definido simplesmente como datas válidas. Sob essa regra, o domínio para *EmployeeHireDate* não inclui 30 de fevereiro de qualquer ano.

## Modelagem e gerenciamento de dados

A Modelagem de Dados é um processo de descoberta e documentação de informações críticas para o entendimento de uma organização por meio de seus dados. Os modelos capturam e permitem o uso do conhecimento dentro de uma organização (ou seja, são uma forma crítica e uma fonte de metadados). Eles podem até ser usados para melhorar a qualidade dessas informações, por meio da aplicação de convenções de nomenclatura e outros padrões que tornam as informações mais consistentes e confiáveis.

Os analistas e designers de dados atuam como intermediários entre os consumidores de informações (as pessoas com requisitos de negócios para dados) e os produtores de dados que capturam os dados de forma utilizável. Os profissionais de dados devem equilibrar os requisitos de dados dos consumidores de informações e os requisitos de aplicação dos produtores de dados.

Os projetistas de dados também devem equilibrar os interesses de curto e de longo prazo. Os consumidores de informações precisam de dados em tempo hábil para atender demandas de curto prazo e avaliar oportunidades. As equipes de projeto de desenvolvimento do sistema devem atender às restrições de tempo e orçamento. No entanto, eles também devem atender aos interesses de longo prazo de todas as partes interessadas, garantindo que os dados de uma organização residam em estruturas de dados seguras, recuperáveis, compartilháveis e reutilizáveis, e que esses dados sejam os íntegros, oportunos, relevantes e utilizáveis.

Portanto, modelos de dados e projetos de banco de dados devem fornecer um equilíbrio razoável entre as necessidades de curto e de longo prazo da organização.

## 3. Habilitando e mantendo dados

Compete a área de arquitetura e modelagem de dados fornecer informações sobre a melhor forma de configurar as aplicações que disponibilizam dados utilizáveis, acessíveis e atualizados para atendimento aos objetivos estratégicos da organização.

### 3.1 Armazenamento e operações de dados

O processo de armazenamento e operação de dados representa o gerenciamento tradicional de dados. Esse ofício é geralmente realizado pelos administradores de banco de dados (Data Base Administrators - DBAs) e pelos Administradores de Armazenamento em Rede (Network Storage Administrators - NSAs), responsáveis por manter o conjunto de dados acessíveis e com a



integridade necessária. O trabalho de armazenamento e operações de dados é essencial para as organizações que dependem de dados para realizar diversos tipos de transações.

A administração do banco de dados é vista como uma função monolítica onde os DBAs desempenham papéis diferentes. Podem suportar ambientes de produção, trabalho de desenvolvimento e procedimentos específicos. O trabalho do DBA é influenciado pelo modelo de Arquitetura de Dados, bem como pela maneira como os próprios bancos de dados são organizados (hierárquicos, relacional, etc.). Com o surgimento de novas tecnologias, os DBAs e NSAs são responsáveis por criar e gerenciar ambientes virtuais (computação em nuvem). Como os ambientes de armazenamento de dados são bastante complexos, os DBAs procuram maneiras de reduzir ou pelo menos gerenciar a complexidade por meio da automação, reutilização e aplicação de padrões e boas práticas reconhecidas.

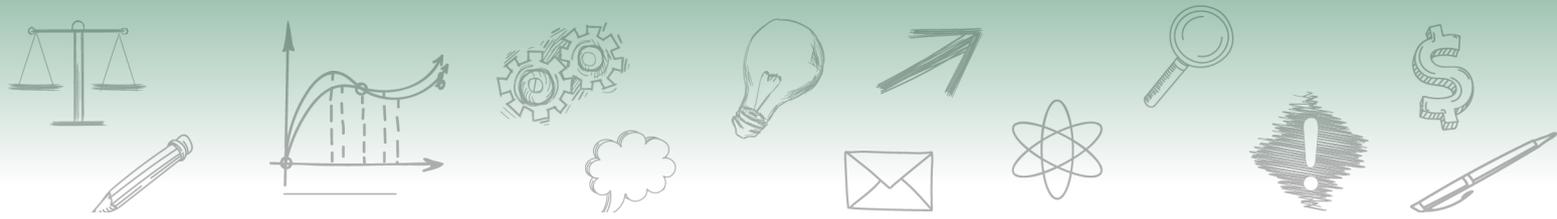
Embora os DBAs possam parecer muito distantes da função de governança de dados, o conhecimento do ambiente técnico é essencial para implementar diretivas de governança relacionada as atividades de controle de acesso, privacidade e segurança dos dados.

O processo de armazenamento e de operação de dados são inerentes ao gerenciamento de dados em todo seu ciclo de vida, desde a coleta até a reutilização ou descarte, conforme a seguir:

- Definir os requisitos de armazenamento.
- Definir os requisitos de acesso.
- Gerenciar o ambiente de armazenamento físico.
- Identificar padrões de uso.
- Planejar a continuidade dos negócios.
- Gerenciar backups.
- Gerenciar a disponibilidade das aplicações.
- Gerenciar os processos de migração de dados.
- Mapear dados como ativos.
- Certificar processos de validação de dados.

## 3.2 Integração e interoperabilidade de dados

Enquanto as atividades de armazenamento e operações de dados se concentram nos ambientes para armazenamento e manutenção, as atividades de integração e interoperabilidade de dados (DII) contemplam os processos para consolidar dados dentro e entre repositórios e aplicações.



A integração trata dos aspectos físicos ou virtuais. Enquanto a Interoperabilidade consiste na capacidade de vários sistemas se comunicarem e interoperar dados de forma eficiente e segura.

As soluções DII permitem funções básicas de gerenciamento de dados, de modo a permitir:

- Migração e conversão de dados.
- Consolidação de dados em Hubs ou Marts.
- Integração de soluções de software de terceiros ao portfólio de aplicações da organização.
- Compartilhamento de dados entre repositórios de dados e datacenters da organização.
- Gerenciamento de interfaces de dados.
- Coleta e integração de dados externos.
- Integração de dados estruturados e não estruturados.
- Fornecimento de inteligência e suporte a decisões de gerenciamento.

A implementação de práticas e soluções de integração e interoperabilidade de dados visa:

- Disponibilizar dados no formato e prazo necessários para as aplicações consumidoras de dados.
- Consolidar dados fisicamente e virtualmente em hubs de dados.
- Reduzir o custo e a complexidade do gerenciamento de soluções, desenvolvendo modelos e interfaces compartilhadas.
- Identificar eventos significativos (oportunidades e ameaças) e acionar alertas e ações automaticamente.
- Apoiar os esforços de *Business Intelligence, analytics e Master Data Management*.

O design das soluções DII deve considerar:

- **Coleta de dados modificados:** como garantir que os dados sejam atualizados corretamente.
- **Latência:** a quantidade de tempo entre quando os dados são coletados ou criados até quando são disponibilizados para consumo.



- **Replicação:** como garantir o desempenho esperado nos processos de compartilhamento de dados.
- **Orquestração:** como diferentes processos são organizados e executados para preservar a consistência e a continuidade dos dados.

O principal fator para o DII é garantir que os dados sejam interoperáveis de forma eficiente para diferentes meios de armazenamento ou aplicações. É relevante ter como objetivo a redução da complexidade nos processos de interoperabilidade.

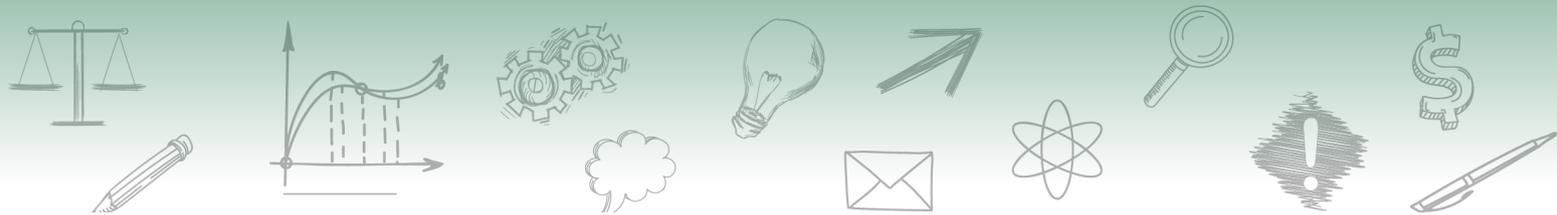
Devido à sua complexidade, o DII depende de outras áreas do gerenciamento de dados para auxiliar na redução de riscos e aumento da eficiência:

- **Governança de dados:** para definir e governar as regras e políticas de transformação de dados e estruturas.
- **Arquitetura de dados:** para projetar soluções.
- **Segurança de dados:** para garantir que as soluções de segurança sejam eficientes nas diversas aplicações de transações de dados entre as organizações.
- **Metadados:** para manter e disponibilizar o inventário de dados (persistente, virtual e em transição), o significado estratégico dos dados, as regras para transformá-los, seu histórico e linhagem.
- **Armazenamento e operações de dados:** para gerenciar as diversas instancias física das soluções.
- **Modelagem e Design de Dados:** para projetar estruturas de dados, incluindo persistência física em bancos de dados, estruturas de dados virtuais e tráfego de informações entre aplicações.

A integração e a interoperabilidade de dados são essenciais para o *Data Warehousing & Business Intelligence*, bem como para o gerenciamento de dados de referência ou de dados mestre, visto que esses dois processos geram transformação e integração de dados de múltiplas origens para hubs de dados consolidados.

A integração e interoperabilidade de dados é ponto central para o gerenciamento de aplicações *Big Data*. O *Big Data* procura integrar diversos tipos de dados, incluindo dados estruturados e dados não estruturados (documentos, arquivos de áudio, vídeo e streaming); que após integrados e tratados podem ser utilizados para desenvolver modelos preditivos e gerar insumos para implantação de atividades de inteligência artificial.

Ao implementar o DII uma organização deve atender os seguintes princípios:



- Adotar uma perspectiva corporativa no design para garantir entregas iterativas e incrementais.
- Equilibrar as demandas por dados compartilhados.
- Garantir a responsabilidade pelo design e atividades do DII envolvidos no design e mudança das regras de transformação de dados, persistentes e virtuais.

O processo de integração refere-se à possibilidade de aplicações serem unificadas, compartilhando informações suficientes para que determinado processo ocorra.

A interoperabilidade pode ser entendida como uma característica que se refere à capacidade de diversos sistemas e organizações trabalharem em conjunto (interoperar) de modo a garantir que pessoas, organizações e sistemas computacionais interajam para trocar informações de maneira eficaz e eficiente.

Outra definição seria que a Interoperabilidade consiste na capacidade de um sistema (informatizado ou não) de se comunicar de forma transparente (ou o mais próximo disso) com outro sistema (semelhante ou não). Para um sistema ser considerado interoperável, é necessário que ele opere com padrões abertos ou ontologias.

## SAIBA MAIS

Hub (traduzido do Inglês, "pivô") ou concentrador é o processo pelo qual se transmite ou difunde determinada informação, tendo, como principal característica, que a mesma informação está sendo enviada para muitos receptores ao mesmo tempo (broadcast). Este termo é utilizado em rádio, telecomunicações e em TI.

Data Mart (repositório de dados) é subconjunto de dados de um Data Warehouse (ou DW, armazém de dados). Geralmente são dados referentes a um assunto ou de diferentes níveis de sumarização (ex: Vendas Anual, Vendas Mensal, Vendas 5 anos), que focalizam uma ou mais áreas específicas. Seus dados são obtidos do DW, desnormalizados e indexados para suportar intensa pesquisa. Data marts extraem e ajustam porções de DWs aos requisitos específicos de grupos/departamentos.

### Importância e Vantagens da Interoperabilidade:

“As principais questões de governança endereçadas pelas instituições abordam cinco itens inter-relacionados: os princípios de TIC; a arquitetura de TIC; a infraestrutura de TIC; as necessidades de aplicações para o negócio; e os investimentos e priorização de TI.” (Governança de Tecnologia da Informação, Peter Weil e Jeanne W. Ross).



A Governança de TIC relacionada à questão da interoperabilidade aborda os “princípios de TIC” e “arquitetura de TI” com o objetivo de definir a abordagem conceitual e tecnológica para interoperar os sistemas de governo, do modo mais eficiente, efetivo e econômico, para o alcance das metas do Governo Federal no atendimento das necessidades da sociedade.

A interoperabilidade gera instrumentos aos gestores para uma melhor tomada de decisão, uma melhor coordenação dos programas e serviços de governo e redução de custos, por meio da visão integrada dos processos do governo e do reuso de soluções tecnológicas.

A interoperabilidade também auxilia no atendimento ao Decreto nº 9.723/2019, que dispõe sobre a simplificação do atendimento prestado aos usuários dos serviços públicos, institui o Cadastro de Pessoas Físicas - CPF como instrumento suficiente e substitutivo para a apresentação de dados do cidadão no exercício de obrigações e direitos e na obtenção de benefícios, ratifica a dispensa do reconhecimento de firma e da autenticação em documentos produzidos no País e institui a Carta de Serviços ao Usuário.”

Os conceitos de interoperabilidade e integração são muitas vezes utilizados como sinônimos na área de TIC. Entretanto, esses conceitos são distintos, embora complementares.

Integração refere-se ao processo de conectar dois ou mais sistemas gerando uma dependência tecnológica entre os mesmos.

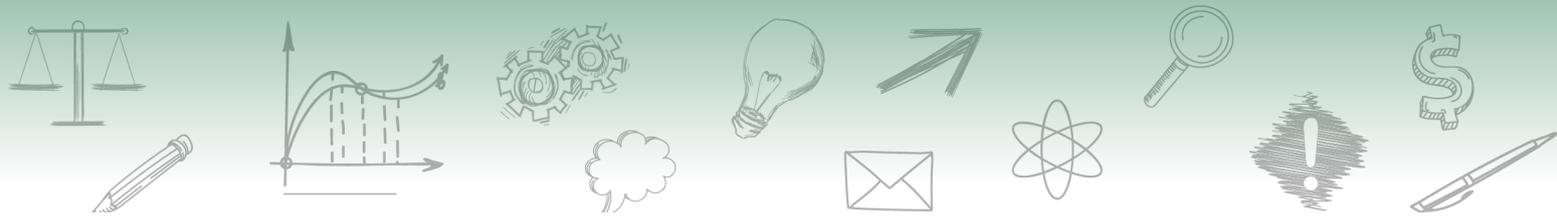
Interoperabilidade refere-se ao processo de comunicação de dois ou mais sistemas sem a geração de uma dependência tecnológica entre os mesmos.

#### Dimensões da Interoperabilidade:

A interoperabilidade pode ser organizada em três dimensões que se comunicam e se complementam: organizacional, semântica e técnica.

A interoperabilidade organizacional diz respeito à colaboração entre organizações que desejam trocar informações mantendo diferentes estruturas internas e processos de negócios variados. Mesmo contando com a padronização de conceitos, as organizações possuem distintos modelos de operação, ou processos de trabalho. Isto quer dizer que elas realizam suas atividades em tempos diferentes e de maneiras diferentes.

A interoperabilidade semântica é a capacidade de dois ou mais sistemas heterogêneos e distribuídos trabalharem em conjunto, compartilhando as informações entre eles com entendimento comum de seu significado (Buranarach, 2004). A interoperabilidade semântica garante que os dados trocados tenham seu efetivo significado corretamente interpretado dentro do



contexto de uma dada transação ou busca de informação, dentro da cultura, convenções e terminologias adotadas por cada setor ou organização e, assim, compartilhados pelas partes envolvidas.

A interoperabilidade técnica trata da ligação entre sistemas e serviços de computação pela utilização de padrões para apresentação, coleta, troca, processamento e transporte de dados. Esses padrões podem abranger hardware, software, protocolos e processos de negócio. Uma vez que foram identificados os motivos e os momentos adequados para interoperar, e que foram estabelecidos vocabulários comuns, é preciso haver também um padrão para fazer isso, ou seja, para tratar o “como fazer”.

Fonte: Manual do Gestor de Interoperabilidade, 2012 - <https://www.governodigital.gov.br/transformacao/compras/orientacoes/interoperabilidade/material-de-apoio>.

O Conecta gov.br é um programa que promove a troca automática e segura de informações entre os sistemas para que o cidadão não tenha que rerepresentar informações que o governo já possua. Isso é um direito do cidadão garantido pela Lei 13.726/2018.

Essa integração de dados, conhecida como interoperabilidade, desonera o cidadão, simplifica o serviço público, reduz fraude e traz segurança e economia para todo o processo.

O Conecta é uma iniciativa da Secretaria de Governo Digital do Ministério da Economia para os órgãos e entidades do Poder Executivo Federal.

#### Benefícios:

Para o cidadão: deixar de preencher formulários, emitir certidões, fazer uploads de documentos e peregrinar entre órgãos para apresentar comprovantes exigidos para a prestação de um serviço público. Isso porque os sistemas do governo conversam entre si e obtém as informações de maneira automática, sem precisar solicitá-las ao cidadão

Para o governo: deixar de conferir manualmente as informações apresentadas pelo cidadão. Também diminui a incidência de erros e fraudes. O atendimento exige menos esforço, fica mais rápido e mais seguro. Isso é economia de dinheiro público.



A existência de uma infraestrutura de Tecnologia da Informação e Comunicação (TICs) estruturada e com fundamentos consolidados para a criação dos serviços de governo digital é o pré-requisito para o fornecimento de melhores serviços à sociedade com custo mais baixos.

Um governo moderno e integrado exige sistemas igualmente modernos e integrados, interoperáveis, trabalhando de forma íntegra, segura e coerente em todo o setor público.

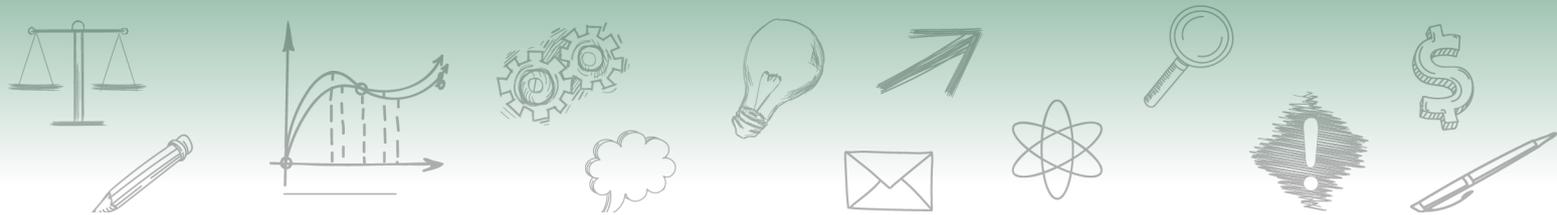
Políticas e especificações claramente definidas para interoperabilidade e gerenciamento das informações são fundamentais para propiciar a conexão do governo, tanto no âmbito interno como na interação com a sociedade e, em maior nível de abrangência, com o resto do mundo, como outros governos e organizações atuantes no mercado mundial.

A arquitetura ePING (Padrões de Interoperabilidade de Governo Eletrônico – Portaria nº 92, de 24 de dezembro de 2014) define um conjunto mínimo de premissas, políticas e especificações técnicas que regulamentam a utilização da Tecnologia de Informação e Comunicação no Governo Federal, estabelecendo as condições de interação com os demais Poderes e esferas de governo e com a sociedade em geral, com mais ênfase nos órgãos e entidades integrantes do Sistema de Administração dos Recursos de Tecnologia da Informação (SISP).

Os recursos de informação do governo constituem valiosos ativos econômicos ao garantir que a informação governamental possa ser rapidamente localizada e transmitida entre os diversos órgãos públicos, mantidas as obrigações de privacidade e segurança. Esse processo é essencial para que o governo utilize da melhor forma esses ativos, melhorando a prestação de serviços públicos no país.

Mais informações: <https://www.gov.br/governodigital/pt-br/governanca-de-dados/padroes-de-interoperabilidade> ou <http://eping.governoeletronico.gov.br/>.





### 3.3 Armazenamento de Dados

Os *data warehouses* (DW) – ou armazém de dados – permitem que as organizações integrem dados de diferentes aplicações em um modelo de dados comum, a fim de suportar funções operacionais, requisitos de conformidade e atividades de *Business Intelligence* (BI). A tecnologia de armazém surgiu nos anos 80. Evoluíram para um viés corporativo a partir dos anos 90. Os DW permitiram que as organizações usassem seus dados com mais eficiência, reduzindo a redundância e proporcionando mais consistência.

O termo armazém de dados implica que todos os dados estão em um único local. No entanto, o conceito de *data warehouses* consiste em várias partes pelas quais os dados são transitados. A estrutura e o formato dos dados podem ser alterados, para que possam ser reunidos em tabelas comuns, a partir das quais podem ser acessados. Pode ser usado diretamente para relatórios ou como entrada para aplicações *downstream*.

A construção de um DW requer conhecimentos de gerenciamento de dados nos processos de operação, integração, de governança de dados. Além disso, é necessário conhecimentos de gerenciamento de processos que permitam que os dados sejam compartilhados com qualidade, de forma segura e utilizável (via metadados confiáveis).

A abordagem adotada no processo de construção de DW dependerá de seus objetivos estratégicos e da arquitetura dos dados. Os DWs compartilham características comuns:

- Armazenam dados de outros sistemas e os tornam acessíveis e utilizáveis para análise.
- O ato de armazenamento inclui organizar os dados de maneira a agregar valor. Em muitos casos, isso significa que os armazéns efetivamente criam novos dados que não estão disponíveis em outros lugares.
- As organizações implantam DWs em razão da necessidade de disponibilizar dados confiáveis e integrados para as partes interessadas autorizadas.
- Os dados armazenados são utilizados para propósitos distintos, principalmente para análise preditiva.

As abordagens mais conhecidas para *data warehousing* foram conduzidas por dois líderes influentes, Bill Inmon e Ralph Kimball.

O Inmon define um *data warehouse* como uma coleta de dados orientada ao assunto, integrada, variável no tempo e não volátil, para apoiar o processo de tomada de decisão da gerência. Um modelo relacional normalizado é usado para armazenar e gerenciar dados. A abaixo ilustra a abordagem da Inmon, que é chamada de "Fábrica de informações corporativas".

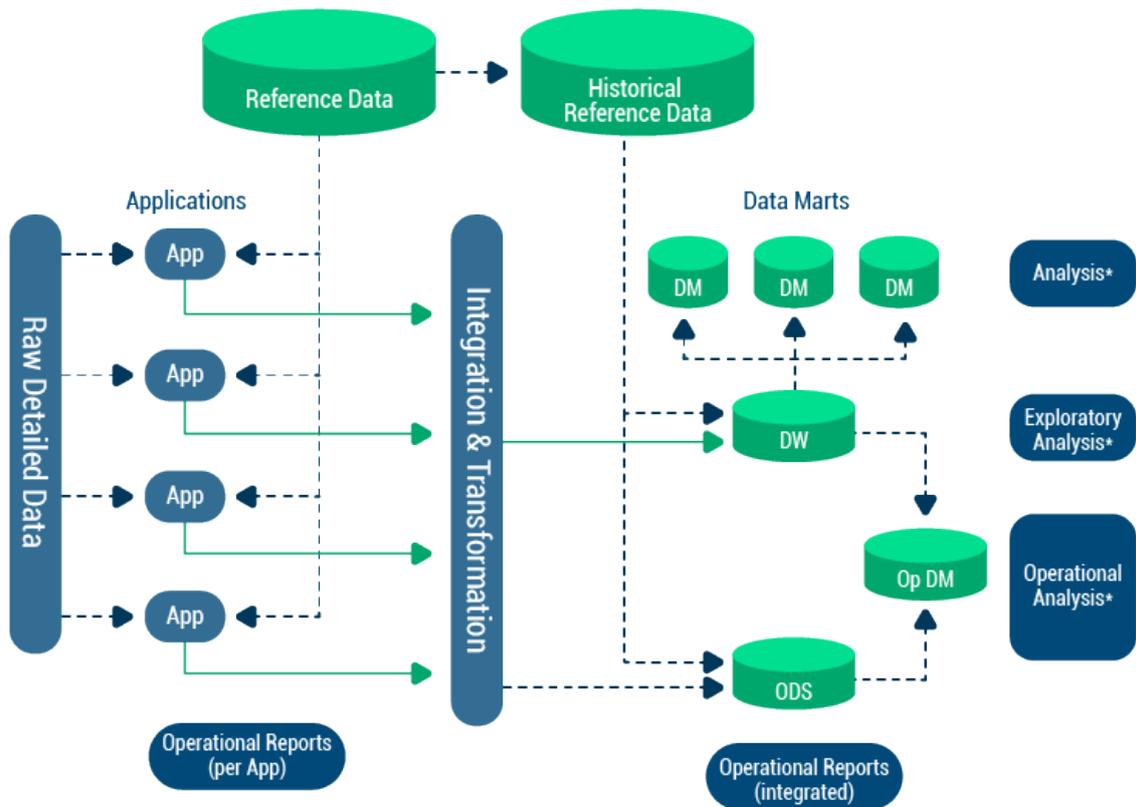
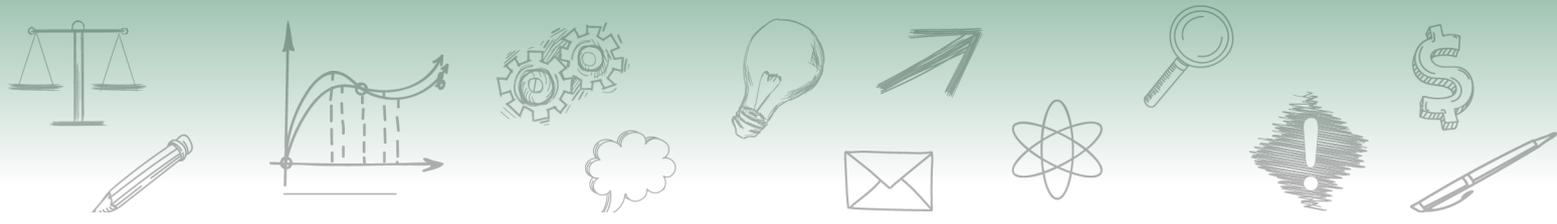


Figura 19: Inmon's Corporate Information Factory (DMBOK2, p. 388)<sup>30</sup>

Tradução dos termos do diagrama (figura 19)

Termo	Tradução
Reference Data	Dados de Referência
Historical Reference Data	Histórico dos Dados de Referência
Applications (App)	Aplicações (App)
Raw Detailed Data	Dados Primários/Brutos Detalhados
Operational Reports (per App)	Relatórios Operacionais
Integrations & Transformation	Integração & Transformação
Data Marts (DM)	Data Marts (DM) – Manter esse termo, que foi explicado neste módulo.
DW	Data Warehouse (DW) - Manter esse termo, que foi explicado neste módulo.
Op. DM	Operação de Dados Mestres (DM)
ODS	Operational data store – Armazenamento de Dados Operacionais
Operational Reports (Integrated)	Relatórios Operacionais (Integração)
Analysis	Camadas Analíticas



Exploratory Analysis	Camada de Análise Exploratória dos Dados
Operational Analysis	Camada de Análise Operacional

Kimball define um *data warehouse* como "uma cópia dos dados de transação estruturados especificamente para consulta e análise". A imagem abaixo ilustra a abordagem de Kimball, que exige um modelo dimensional.

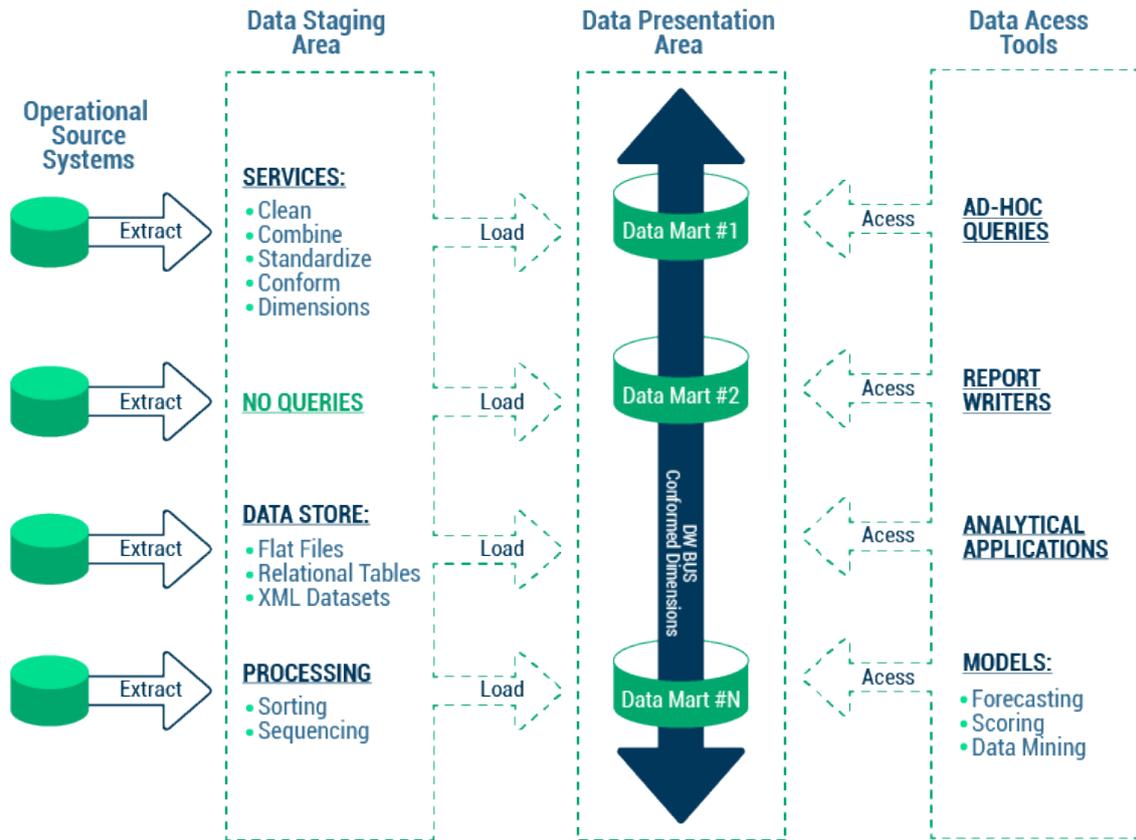


Figura 20: Kimball's Data Warehouse Chess Pieces (DMBOK2, p. 390)<sup>31</sup>

Tradução - Termos do diagrama (figura 20)

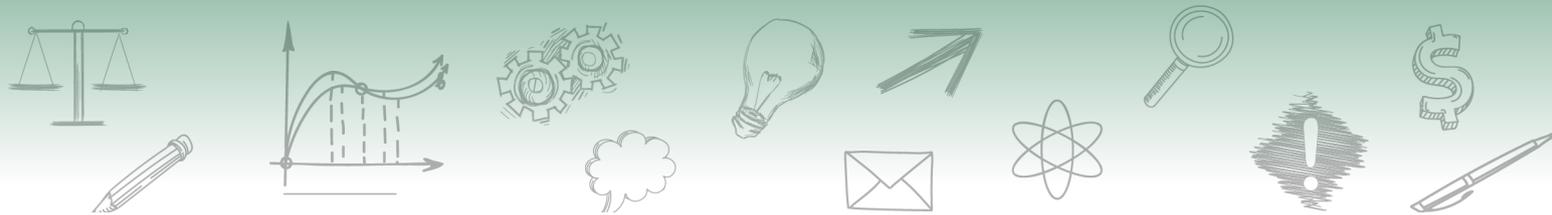


Termo	Tradução
Data Staging Area	Local de Armazenamento dos Dados
Operational Source Systems <ul style="list-style-type: none"> <li>• Extract</li> <li>• Load</li> <li>• Services               <ol style="list-style-type: none"> <li>a. Clean</li> <li>b. Combine</li> <li>c. Standardize</li> <li>d. Conform</li> <li>e. Dimensions</li> </ol> </li> <li>• No Queries</li> <li>• Data Store               <ol style="list-style-type: none"> <li>a. Flat Files</li> <li>b. Relational Tables</li> <li>c. XML Datasets</li> </ol> </li> <li>• Processing               <ol style="list-style-type: none"> <li>a. Sorting</li> <li>b. Sequencing</li> </ol> </li> </ul>	Sistemas de Origem Operacional <ul style="list-style-type: none"> <li>• Extração</li> <li>• Carga</li> <li>• Serviços               <ol style="list-style-type: none"> <li>a. Limpeza</li> <li>b. Combinação/intercâmbio</li> <li>c. Padronização</li> <li>d. Conformidade</li> <li>e. Dimensionamento</li> </ol> </li> <li>• Sem Consultas</li> <li>• Armazenamento de Dados               <ol style="list-style-type: none"> <li>a. Arquivos Simples/Uniformes</li> <li>b. Tabelas de Relacionamento</li> <li>c. Conjunto de Dados em XML</li> </ol> </li> <li>• Processamento               <ol style="list-style-type: none"> <li>a. Ordenação</li> <li>b. Sequenciamento</li> </ol> </li> </ul>
Data Presentation Area	Local de Apresentação dos Dados
<ul style="list-style-type: none"> <li>• Data Mart</li> <li>• DW BUS - Conformed Dimensions</li> </ul>	<ul style="list-style-type: none"> <li>• Data Mart</li> <li>• Arquitetura de Barramento de DW - Dimensões em Conformidade</li> </ul>
Data Access Tools	Ferramentas de Acesso aos Dados
<ul style="list-style-type: none"> <li>• Ad Hoc Queries</li> <li>• Report Writers</li> <li>• Analytical Applications</li> <li>• Models               <ol style="list-style-type: none"> <li>a. Forecasting</li> <li>b. Scoring</li> <li>c. Data Mining</li> </ol> </li> </ul>	<ul style="list-style-type: none"> <li>• Consultas Ad Hoc (“Para esta finalidade”)</li> <li>• Gerar Relatórios</li> <li>• Aplicações Analíticas</li> <li>• Modelos               <ol style="list-style-type: none"> <li>a. Análises Preditivas</li> <li>b. Indicadores</li> <li>c. Mineração de Dados</li> </ol> </li> </ul>

As organizações estão construindo DWs de segunda e terceira geração ou adotando *Data Lakes* (lagos de dados) para disponibilizar mais dados a uma velocidade mais elevada, criando a oportunidade de aprimorar as análises preditivas.

### 3.4 Gerenciamento de Dados de referência

Diferentes tipos de dados desempenham diferentes funções dentro de uma organização e têm diferentes requisitos de gerenciamento de dados. Dados de referência (por exemplo, tabelas de código e descrição) são dados usados apenas para caracterizar outros dados em uma organização



ou apenas para relacionar dados em um banco de dados a informações além dos limites da organização.

Os Dados d

]~]~]e referência fornecem um contexto crítico para a criação e uso de dados mestre e transacionais. Permite que outros dados sejam compreendidos de maneira significativa. Ter várias instâncias dos mesmos Dados de Referência é ineficiente e inevitavelmente resulta em inconsistência. A inconsistência gera ambiguidades, a qual produz riscos.

O gerenciamento de dados de referência (*Reference Data Management - RDM*) envolve controle sobre dados de domínio, ou seja, garantir que a organização tenha acesso a um conjunto de dados de referência que sejam precisos e atuais para cada conceito representado.

A propriedade e a responsabilidade pelos Dados de Referência são um desafio para as organizações. Alguns dados de referência se originam fora da organização, outros podem ser gerados e mantidos internamente. Determinar a responsabilidade pela gestão dos dados de referência faz parte do gerenciamento. Sendo que a falta dessa responsabilização gera riscos, pois as inconsistências nos Dados de referência podem causar prejuízos irreparáveis no contexto da gestão de dados.

O conjunto de dados de referência geralmente são menores e menos voláteis do que outras formas de dados.

O gerenciamento de dados de referência torna-se eficaz (atualizados e consistentes em várias aplicações) quando realizado por meio de tecnologias que permitam aos consumidores e as partes interessadas acessá-los de maneira oportuna e eficiente, independentemente da plataforma adotada.

O gerenciamento de dados de referência requer planejamento e design. A arquitetura e os modelos de dados de referência devem considerar que os dados de referência serão armazenados, mantidos e compartilhados. Com isso, há necessidade de estabelecer políticas de segurança mais rígidas.

### 3.5 Gerenciamento de Dados Mestre

Gerenciamento de Dados Mestres (*Master Data Management - MDM*) é uma estrutura de processos e tecnologias destinadas a criar e manter um ambiente de dados autorizado, confiável e sustentável, preciso e seguro que representa uma versão única da verdade, um sistema de registro aceito usado tanto intra - e entre organizações em um conjunto diversificado de sistemas e aplicações, estratégias de negócio e comunidades de usuários. (*Alex Berson e Larry Dubov "Gerenciamento de dados mestre e integração de dados do cliente para uma empresa global" McGraw Hill, 2007*).



O gerenciamento de dados mestre é um desafio, visto que dados são alterados de forma tempestiva, exigindo políticas e normas desenhadas para essa finalidade.

O planejamento para o gerenciamento de dados mestre inclui várias etapas:

- Identificar fontes de dados elegíveis e que atendam aos requisitos que justificam sua escolha como Dados Mestres.
- Elaborar políticas e regras específicas.
- Estabelecer regras para identificar e tratar dados redundantes.
- Estabelecer regras para compartilhamento de dados confiáveis, sustentáveis, precisos e seguros.

O MDM é um processo de gerenciamento do ciclo de vida dos dados. Não deve ser gerenciado apenas pelo sistema MDM, mas compartilhado para uso em outras aplicações. Isso requer tecnologia que permita compartilhamento e feedback. Todo esse processo deve estar alinhado à políticas específicas para MDM, de forma que as regras sejam internalizadas nos processos de negócio para uso de dados oriundos de um processo de qualidade de dados com a finalidade de evitar inconsistências.

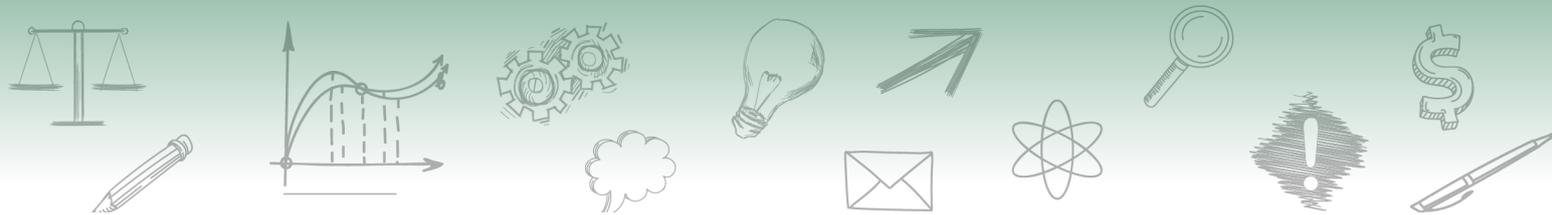
Os Dados Mestres gerenciados de forma eficiente, ampliam a eficiência organizacional e reduzem os riscos associados a inconsistências na estrutura de dados entre sistemas e processos.

### 3.6 Gerenciamento de documentos e conteúdo

Documentos, registros e conteúdo (por exemplo, as informações armazenadas nos sites da Internet e da intranet) compreendem uma forma de dados com requisitos de gerenciamento distintos. O Gerenciamento de documentos e conteúdo envolve controlar a coleta, armazenamento, o acesso e uso de dados e informações compartilhados ou não. Dados não estruturados requer um tratamento que garanta sua confiabilidade e qualidade. Para isso é necessária uma estrutura de governança, arquitetura confiável e metadados bem gerenciados.

O gerenciamento de documentos e de conteúdo se concentram em manter a integridade e permitir o acesso a documentos e outras informações não estruturadas ou semiestruturadas. No entanto, também possui direcionadores estratégicos. Os principais fatores de negócios para gerenciamento de documentos e conteúdo incluem conformidade regulamentar, capacidade de responder a solicitações de litígio e descoberta eletrônica e requisitos de continuidade de negócios.

O gerenciamento de documentos é o termo usado para descrever armazenamento, inventário e controle de documentos eletrônicos ou em meio físico. Abrange as técnicas e tecnologias para controlar e organizar documentos ao longo de seu ciclo de vida.



O Gerenciamento de Registros é uma forma especializada de gerenciamento de documentos que se concentra nos registros - documentos que fornecem evidências das atividades de uma organização. Essas atividades podem ser eventos, transações, contratos, correspondência, políticas, decisões, procedimentos, operações, arquivos de pessoal, etc. Os registros podem ser documentos físicos, arquivos e mensagens eletrônicas ou conjunto de dados.

O Gerenciamento de conteúdo refere-se aos processos, técnicas e tecnologias para organizar, categorizar e estruturar recursos de informações para que possam ser armazenados, publicados e reutilizados de várias maneiras. O conteúdo pode ser volátil ou estático (documentos e outros ativos digitais, como vídeos, fotografias, etc.). Pode ser gerenciado formalmente (estritamente armazenado, gerenciado, auditado, retido ou descartado) ou informalmente por meio de atualizações ad hoc. O gerenciamento de conteúdo é particularmente importante em sites e portais, mas as técnicas de indexação com base em palavras-chave e organização com base em taxonomias podem ser aplicadas em plataformas tecnológicas.

O gerenciamento eficiente de documentos, registros e outras formas de conteúdo compartilhado requer:

- Planejamento, incluindo a criação de políticas para diferentes tipos de acesso e manipulação.
- Definição da arquitetura da informação e dos metadados necessários para suportar uma estratégia de conteúdo.
- Permissão do gerenciamento de terminologia, incluindo ontologias e taxonomias necessárias para organizar, armazenar e recuperar várias formas de conteúdo.
- Adoção de tecnologias que permitem o gerenciamento do ciclo de vida do conteúdo, desde a criação ou captura de conteúdo até a versão e garantido a segurança do conteúdo.

Para os registros as políticas de retenção e descarte são críticas. Os registros devem ser mantidos pelo período de tempo necessário e devem ser destruídos assim que os requisitos de retenção forem atendidos. Embora existam, os registros devem estar acessíveis aos usuários e a processos apropriados e, como outros conteúdos, devem ser compartilhados por meio de canais apropriados.

Atingir esses objetivos necessita de sistemas de gerenciamento de conteúdo (Content Management Systems - CMS), além de ferramentas para criar e gerenciar os metadados que suportam o uso do conteúdo. A governança de dados deve elaborar as políticas e procedimentos que suportam o uso de conteúdo e evitam o uso indevido; permitindo que a organização responda a qualquer forma de litígio de maneira consistente e apropriada.



### 3.7 Big Data

O *big data* e a ciência de dados estão conectados a mudanças tecnológicas significativas que permitiram aos usuários gerar, armazenar e analisar quantidades cada vez maiores de dados e usá-los para prever e influenciar comportamentos, além de obter informações sobre uma série de assuntos como práticas de saúde, gestão de recursos naturais e desenvolvimento econômico.

Os primeiros esforços para definir o significado de *Big Data* caracterizaram-no em termos dos Três V's: Volume, Velocidade, Variedade. À medida que mais organizações começam a aproveitar o potencial do *Big Data*, a lista de V's se expandiu:

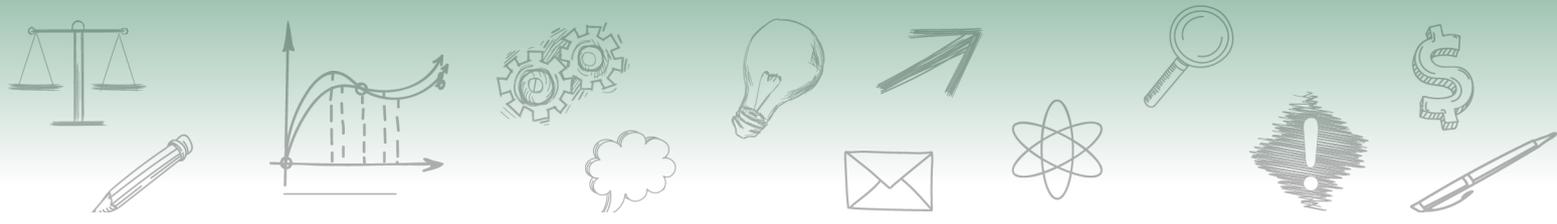
- **Volume:** refere-se à quantidade de dados. O *Big Data* geralmente possui milhares de entidades ou elementos em bilhões de registros.
- **Velocidade:** refere-se à velocidade na qual os dados são capturados, gerados ou compartilhados. O *Big Data* geralmente é gerado e também pode ser distribuído e até analisado em tempo real.
- **Variedade/variabilidade:** refere-se às formas em que os dados são capturados ou entregues. O *Big Data* vem requer armazenamento de vários formatos. A estrutura de dados geralmente é inconsistente dentro ou entre conjuntos de dados.
- **Viscosidade:** refere-se à dificuldade em usar ou integrar os dados.
- **Volatilidade:** refere-se à frequência com que os dados são alterados e, portanto, quanto tempo os dados são úteis.
- **Veracidade:** refere-se à confiabilidade dos dados.

Tirar proveito do *Big Data* requer mudanças na tecnologia e nos processos de negócio e na maneira como os dados são gerenciados. A maioria dos data warehouses - DW é baseado em modelos relacionais. O DW depende do conceito de ETL (*loading and then transforming* - extrair, transformar, carregar). As soluções de *Big Data*, como os lagos de dados, dependem do conceito de E L T - carregar e depois transformar (*loading and then transforming*).

A velocidade e o volume dos dados apresentam desafios que exigem abordagens diferentes para aspectos críticos do gerenciamento de dados, não apenas a integração, mas também o gerenciamento de metadados, a avaliação da qualidade e o armazenamento de dados.

A capacidade de gerenciar o *Big Data* requer mais disciplina do que o gerenciamento de dados. Cada um dos V's apresenta a oportunidade para o caos.

O *Big Data* pode ser usado para uma série de atividades, de mineração de dados, aprendizado de máquina e análise preditiva. Mas para isso é necessário criar uma estratégia para implementação do *Big Data* alinhada a estratégia de negócio da organização.



Deve-se avaliar:

- **Para que a organização precisa de análise:** Uma organização pode determinar que os dados devem ser usados para avaliar suas estratégias de negócio.
- **Quais fontes de dados:** As fontes internas podem ser acessíveis, mas podem ter escopo limitado. Fontes externas podem ser úteis, mas estão fora do domínio da Governança de Dados.
- **Escopo dos dados:** muitos elementos podem ser fornecidos em feeds em tempo real, instantâneos em um determinado momento ou até integrados. Dados de baixa latência são ideais, mas costumam custar recursos de aprendizado de máquina.
- **O impacto e a relação com outras estruturas de dados:** pode ser necessário fazer alterações na estrutura ou no conteúdo de outras estruturas de dados para torná-los adequados para integração com o *Big Data*.

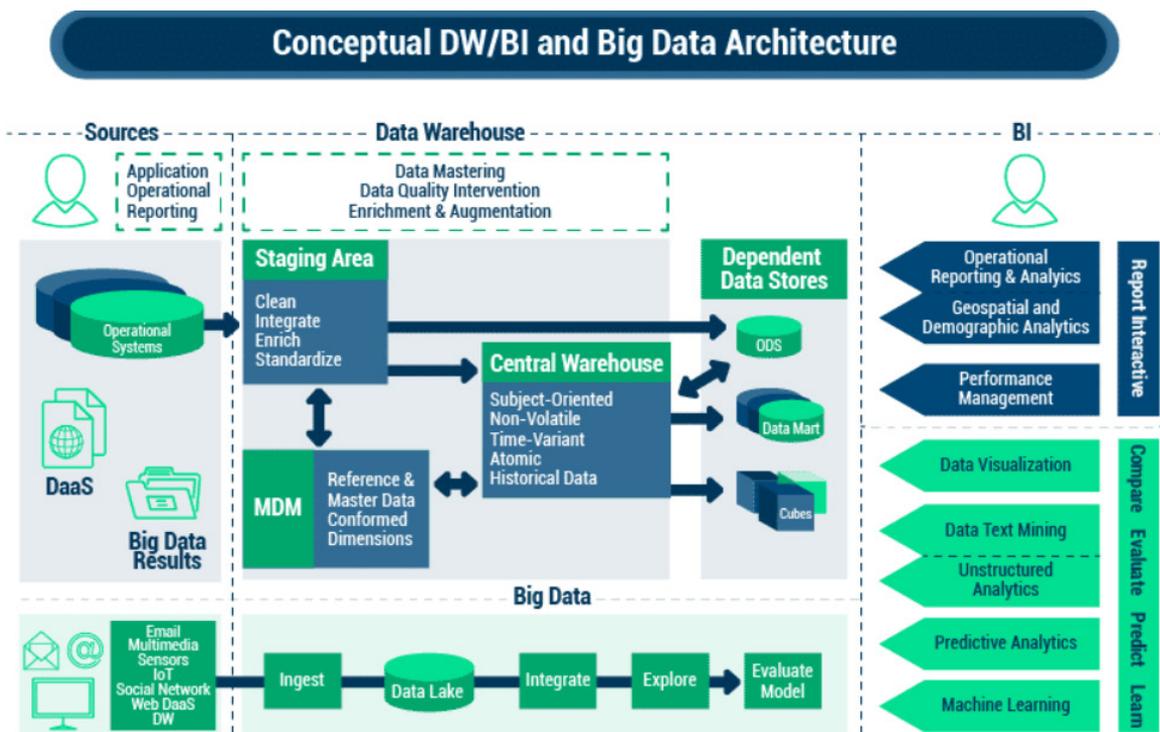


Figura 21: Conceptual DW/BI and Big Data Architecture



## 4. Usando e aprimorando dados

Aplicações utilizam os mesmos dados de forma simultânea. Os dados não se esgotam como também o seu uso geram mais dados. Por exemplo, agregações e cálculos de conjuntos de dados existentes criam novos conjuntos de dados, assim como modelos preditivos criados por cientistas de dados. Esses novos conjuntos de dados continuam sendo produzidos e atualizados. Neste caso, faz necessário a implantação de estruturas com a finalidade de gerenciar por meio de metadados, bem como suportar os processos de qualidade e de controles de segurança.

### 4.1 Inteligência do Negócio

O desenvolvimento de relatórios de *Business Intelligence* é outra atividade em que o uso de dados resulta na criação de novos dados que exige um nível de gerenciamento contínuo.

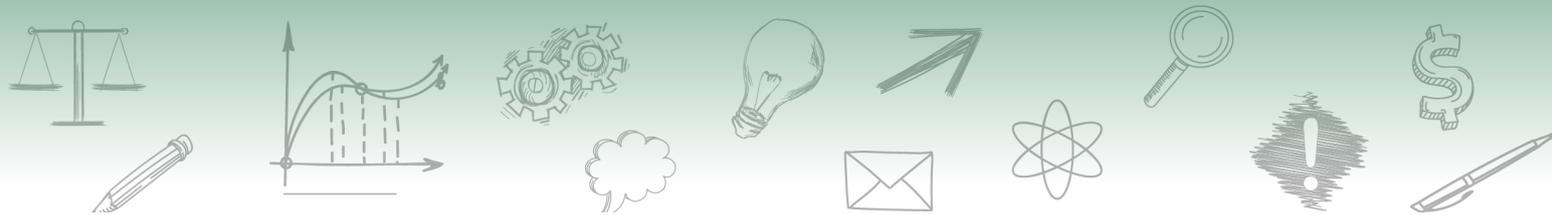
O termo *Business Intelligence* (BI) é conceituado de duas formas:

- **Primeiro**, refere-se a um **tipo de análise de dados** que visa entender as atividades e oportunidades organizacionais, por meio de insights extraídos de análises dos dados locais e exportados, permitindo apoiar a tomada de decisão em relação ao alcance dos seus objetivos estratégicos.
- Segundo, *Business Intelligence* refere-se a **um conjunto de tecnologias que suportam processos de análise de dados**, permitindo realizar consultas, mineração de dados, análise estatística, geração de relatórios, modelagem de cenários, visualização de dados e criação de painéis de controle e de acompanhamento.

As atividades tradicionais de BI exigem fontes de dados confiáveis, integradas e disponíveis. As ferramentas de BI devem suportar a exploração de dados e deve se orientar por processos, de forma a:

- Manter e aprimorar os dados estratégicos utilizados nos relatórios de BI e permitindo a incorporação de novos dados.
- Manter e aprimorar o conjunto de ferramentas de BI.
- Gerenciar metadados para que otimizar relatórios elegíveis.
- Documentar a linhagem de dados para as partes interessadas.
- Fornecer um ciclo de feedback da qualidade dos dados que possibilite identificar oportunidades e aprimoramentos.

O gerenciamento dos dados implementado por um programa de BI segue as mesmas etapas do gerenciamento do ciclo de vida dados.



## 4.2 Ciência de Dados

A ciência de dados existe há muito tempo. Era identificada como Estatística Aplicada. Mas a capacidade de explorar padrões de dados evoluiu rapidamente no século XXI com o advento das tecnologias de coleta e armazenamento de dados, a exemplo dos *Datas Warehouses* (DW) e *Big Data*.

A ciência de dados mescla mineração de dados, análise estatística e aprendizado de máquina com recursos de integração e modelagem de dados para se criar modelos preditivos que exploram padrões de conteúdo de dados. O termo ciência de dados refere-se ao processo de desenvolvimento de modelos preditivos. O analista de dados (ou cientista de dados) usa o método científico (observação, hipótese, experimentação, análise e conclusão) para desenvolver e avaliar um modelo analítico ou preditivo.

O cientista de dados desenvolve uma hipótese sobre o comportamento que pode ser observado nos dados antes de uma ação específica. Por exemplo, a compra de um tipo de item geralmente é seguida pela compra de outro tipo de item. Em seguida, o cientista de dados analisa grandes volumes de dados históricos para determinar com que frequência a hipótese foi verdadeira no passado e para verificar estatisticamente a precisão provável do modelo.

Se uma hipótese é válida com frequência suficiente e se o comportamento previsto é útil, o modelo pode se tornar a base de um processo de inteligência operacional para prever o comportamento futuro, possivelmente até mesmo em tempo real.

De certa forma, a ciência de dados pode ser entendida como uma extensão do BI. O *Business Intelligence* tradicional fornece relatórios do 'passado' - análise de dados estruturados para descrever tendências passadas. Em alguns casos, os padrões de BI são utilizados para prever o comportamento futuro, mas com ausência de precisão confiável.

## 4.3 Análise preditiva e prescritiva

A ciência de dados está focada em criar modelos preditivos. A forma mais simples de modelo preditivo é a previsão. O *Predictive Analytics* é o subcampo do aprendizado de máquina supervisionado, enraizado nas estatísticas em que os usuários tentam modelar elementos de dados e prever resultados futuros por meio da avaliação de estimativas de probabilidade.

O *Predictive Analytics* utiliza modelos de probabilidade com base em variáveis (incluindo dados históricos) relacionadas a possíveis eventos. Quando recebe outras informações, o modelo desencadeia uma reação da organização. O fator desencadeante pode ser um evento como um *feed* de notícias ou dados de sensores de utilidade ou um aumento no volume de solicitações de serviço. O fator de disparo pode ser um evento externo. As notícias relatadas sobre uma organização podem servir como um preditor de uma mudança no preço das ações.

O fator desencadeante é o acúmulo de grande volume de dados em tempo real decorrente do elevado número de negociações, solicitações de serviço e volatilidade do ambiente.



O tempo de resposta de um modelo preditivo, que é calculado entre a previsão e a ocorrência do evento previsto é bastante reduzido (em segundos). O investimento em soluções de tecnologia de latência, como bancos de dados em memória, redes de alta velocidade otimiza a capacidade da organização em ampliar a eficiência das previsões.

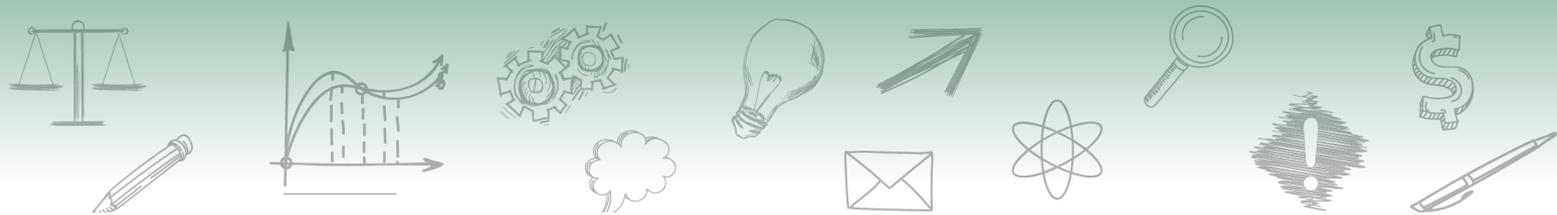
A análise prescritiva tende a ser uma análise preditiva, mas com novas abordagens estatísticas que possibilite prever resultados, e não apenas avaliar os resultados de eventos pretéritos. A análise prescritiva antecipa o que irá ocorrer, quando acontecerá e implica por que isso irá acontecer. Como a análise prescritiva pode mostrar as implicações de várias decisões, pode sugerir como aproveitar uma oportunidade ou evitar um risco. A análise prescritiva pode captar continuamente novos dados para submetê-los a um novo processo de análise prescritiva. Esse processo amplia a qualidade da previsão resultando em prescrições aprimoradas. A tabela a seguir resume a relação entre BI tradicional e Ciência de Dados.

Data Warehouse / BI Tradicional	Ciência de Dados / Análise Preditiva	Ciência de Dados / Análise Prescritiva
Retrospectiva	Percepção	Predição
Baseado na história: O que aconteceu? Por que isso aconteceu?	Com base em modelos preditivos: O que é provável que aconteça?	Com base em cenários: O que devemos fazer para que os eventos ocorram?
<b>Descritivo</b>	<b>Preditivo</b>	<b>Prescritivo</b>

Até recentemente, a análise de grandes conjuntos de dados era limitada pela tecnologia. As análises basearam-se em amostragem ou por outros meios de abstração para aproximar padrões. À medida que cresce a capacidade de coletar e analisar grandes conjuntos de dados, os cientistas de dados estão integrando métodos de matemática, estatística, ciência da computação, processamento de sinais, modelagem de probabilidade, reconhecimento de padrões, aprendizado de máquina, modelagem de incertezas e visualização de dados para obter insights e prever comportamentos com base em conjuntos de *Big Data*. Em suma, a ciência de dados encontrou novas maneiras de analisar e extrair conhecimento dos dados. Em muitos casos, esse conhecimento pode ser traduzido em valor econômico.

Como o *Big Data* foi introduzido nos ambientes de data warehouse e BI, as técnicas de ciência de dados podem fornecer uma visão prospectiva da organização. Recursos preditivos, em tempo real e baseados em modelo, usando diferentes tipos de fontes de dados, oferecem às organizações uma melhor visão para onde estão caminhando.

Os modelos de ciência de dados se tornam fontes de dados, que precisam ser monitorados e explorados para obter informações. Como outras formas de ciência, a ciência de dados cria novos conhecimentos e também novas hipóteses. Testar hipóteses resulta em novos modelos e novos dados. Todas essas peças requerem gerenciamento para se criar o valor ao longo do tempo. Os



modelos precisam ser 'treinados' e avaliados. Novas fontes de dados podem ser incorporadas aos modelos existentes. O ciclo de vida dos dados deve compor os esforços de ciência de dados que precisa ser contabilizado como parte do planejamento e da estratégia.

## DICA

Os modelos de ciência de dados que contêm lógica (algoritmos) para processar dados e fazer previsões a partir deles, não são os mesmos modelos de dados descritos no tópico “Habilitando e Mantendo Dados” que documentam a estrutura dos dados e os relacionamentos entre entidades e atributos de dados.

### 4.4 Visualização de dados

Visualização é o processo de interpretação de conceitos, ideias e fatos utilizando figuras ou representações gráficas. A visualização de dados facilita a compreensão dos dados subjacentes, resumindo-os em uma forma visual, como um gráfico. As visualizações condensam e encapsulam dados e características facilitando sua visualização. Dessa forma podem revelar oportunidades, identificar riscos ou destacar novas análises.

A visualização tem sido fundamental para a análise de dados. As ferramentas tradicionais de BI incluem opções de visualização, como tabelas, gráficos de diversos modelos, histogramas, dentre outras formas.

Em aplicações de ciência de dados, a visualização de dados é fundamental porque, sem ela, a interpretação dos dados é quase impossível. Padrões identificados em um grande conjunto de dados podem ser difíceis, se não impossíveis, de reconhecer em uma exibição de números. Um padrão pode ser capturado rapidamente, quando milhares de pontos de dados são apresentados em uma exibição visual.

As visualizações de dados podem ser entregues em um formato estático, como um relatório publicado ou em um formato online mais interativo. Algumas tecnologias de visualização permitem que os dados se movam entre as camadas de dados, através de filtros ou a capacidade de aprofundar os dados. Outros permitem que a visualização seja alterada pelo usuário sob demanda através de displays inovadores, como mapas de dados e paisagens em movimento de dados ao longo do tempo.

O número de ferramentas de visualização aumentou e as técnicas melhoraram. À medida que a análise de dados evolui, a visualização apresenta novas maneiras de exibição oferecendo vantagens estratégicas. Ver novos padrões nos dados pode resultar em novas oportunidades de negócio. À medida que a visualização de dados continua a evoluir, as organizações terão que ampliar suas equipes de *Business Intelligence* para competir em um mundo cada vez mais orientado a dados.



Um fator crítico de sucesso na implementação de uma abordagem de ciência de dados é o alinhamento das ferramentas de visualização apropriadas à comunidade de usuários. Dependendo do tamanho e da natureza da organização, é provável que muitas ferramentas de visualização diferentes sejam aplicadas em uma variedade de processos. Certifique-se de que os usuários entendam a complexidade relativa das ferramentas de visualização. Usuários sofisticados terão demandas cada vez mais complexas. A coordenação entre as equipes de arquitetura corporativa, gerenciamento de portfólio e manutenção será necessária para controlar os canais de visualização dentro e através do portfólio. Esteja ciente de que a mudança de provedores de dados ou critérios de seleção provavelmente terá impactos nos elementos disponíveis para visualização que podem afetar a eficácia das ferramentas.

É uma prática recomendada estabelecer uma comunidade que defina e publique padrões e diretrizes de visualização e revise artefatos dentro de um método de entrega especificado; isso é particularmente vital para o conteúdo voltado para clientes e regulamentações.

A visualização de dados cria novos conjuntos de dados, na forma das próprias visualizações e nos métodos pelos quais os dados são combinados para que possam ser apresentados em um formato gráfico. Esses dados também devem ser gerenciados.

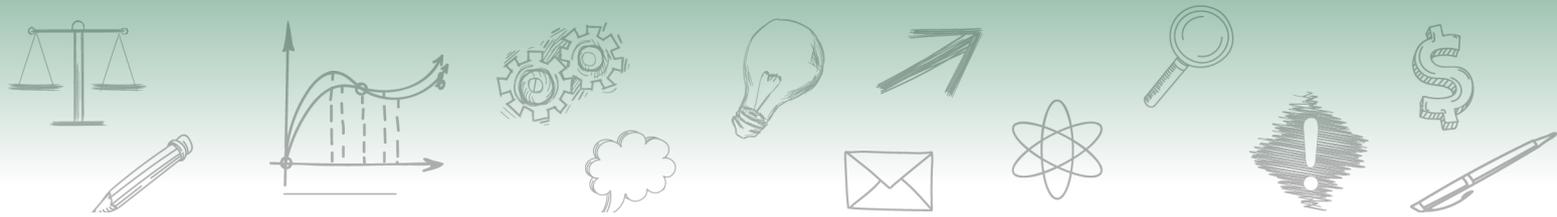
## DICA

A visualização de dados é um campo em evolução, mas os princípios que os guiam se baseiam nos princípios de design. Ver Tufte (2001) e McCandless (2012). Existem vários recursos baseados na Web com exemplos e exemplos contrários. Consulte a Tabela Periódica de Métodos de Visualização em Visual Literacy.Org <https://bit.ly/IX1bvl>.

## 4.5 Monetização de Dados

É provável que qualquer organização envolvida na Data Science ou outras formas de análise obtenha informações valiosas sobre seus próprios clientes, produtos, serviços e processos. A análise avançada também pode gerar informações sobre entidades externas. É provável que essa organização desenvolva técnicas que possam ser relevantes para outros usuários. Em alguns círculos, a monetização direta de dados é percebida como o santo graal do gerenciamento de dados. Algumas organizações (Dun & Bradstreet, Google, Amazon) fizeram um negócio de monetizar seus dados. Mas vender dados e informações não é a única maneira de obter valor dos ativos de dados.

Em *Monetizing Data Management*, Peter Aiken e Juanita Billings, os dois autores apontam que poucas organizações exploram a vantagem estratégica que podem obter com os dados, “o único ativo estratégico não esgotável, não degradante, durável e duradouro da organização”. Eles argumentam que melhorar as práticas de gerenciamento de dados é o primeiro meio de obter mais valor dos dados. Uma organização que coloca um valor monetário em práticas eficazes de gerenciamento de dados produzirá dados de maior qualidade e poderá fazer mais com eles.



Aiken e Billings afirmam que boas práticas de gerenciamento de dados também são a base para a inovação bem-sucedida do uso de dados. As práticas precárias de gerenciamento de dados, por outro lado, custam dinheiro e introduzem riscos em novas iniciativas e processos existentes. Os autores apresentam estudos de caso que documentam que más práticas de gerenciamento de dados podem resultar em desperdício direto por meio de trabalho redundante e, com ele, a criação de dados redundantes, metadados ruins ou ausentes, processos confusos e informações incorretas. Eles também fornecem exemplos dos benefícios de práticas disciplinadas de gerenciamento de dados. Por exemplo, práticas claras e executáveis de gerenciamento de metadados aumentam o conhecimento organizacional e tornam esse conhecimento transferível.

*Infonomics*, de Douglas Laney, um estudo completo sobre gerenciamento de informações como um ativo, apresenta uma ampla variedade de estudos de caso que demonstram como as organizações aproveitaram seus ativos de informações para criar valor. Enquanto as indústrias, atividades e produtos diferem, o valor econômico derivado dos dados se resume a dois métodos básicos:

1. Troca de informações por bens, serviços ou dinheiro.
2. Usando informações para aumentar a receita, reduzir despesas e gerenciar riscos.

Laney apresenta 12 (doze) drivers (direcionadores) de negócio para monetizar dados. Uma das primeiras maneiras de se obter valor é usar os dados organizacionais com mais eficiência para reter, entrar em novos mercados e criar novos produtos. Mas Laney vai além do óbvio. Por exemplo, dados de qualidade podem melhorar a eficiência organizacional, permitindo que uma organização reduza os custos de manutenção, negocie melhores termos e condições, detecte fraudes e desperdícios ou se prepare para assumir os custos de gerenciamento de dados.

Os estudos de caso de Aiken e Billings e Laney e de outras pesquisas confirmam que dados de baixa qualidade são um passivo significativo. Outros, no entanto, conseguiram avançar, com melhorias operacionais e monetização direta. Os estudos de caso mostram que o uso inovador de dados requer gerenciamento confiável de dados. Embora nem toda organização almeje vender seus dados, todas as organizações vislumbram ter confiança nas decisões com base em seus dados. O primeiro passo nessa direção é gerenciar bem os dados.

## 5. Proteção de Dados, Privacidade, Segurança e Gerenciamento de Riscos

O gerenciamento de dados por meio do seu ciclo de vida depende de um conjunto de processos fundamentais que permite o uso e aprimoramento contínuos dos dados. Isso inclui a proteção de dados contra o uso não autorizado, o gerenciamento de metadados (o conhecimento necessário para entender e usar dados) e o gerenciamento da qualidade dos dados. Conforme observado anteriormente, as atividades fundamentais devem ser consideradas como parte do planejamento e design e devem ser realizadas operacionalmente. Essas atividades também são apoiadas pelas estruturas de governança.

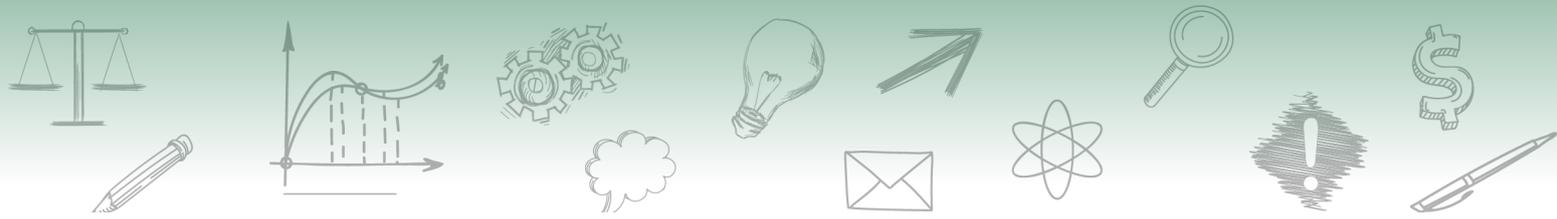


Este tópico abordará a proteção e segurança dos dados. A Segurança de dados inclui o planejamento, desenvolvimento e execução de políticas e procedimentos de segurança para fornecer autenticação, autorização, acesso e auditoria adequados aos ativos de dados e informações.

## 5.1 Objetivos de segurança de dados

As especificidades da segurança de dados (que dados precisam ser protegidos, por exemplo) diferem entre indústrias e países. Mas o objetivo das práticas de segurança de dados é o mesmo: proteger os ativos de informações em alinhamento com os regulamentos de privacidade e confidencialidade, acordos contratuais e requisitos de negócios. Esses requisitos vêm de:

- **Partes interessadas**  
As organizações devem reconhecer as necessidades de privacidade e confidencialidade de suas partes interessadas, incluindo clientes, pacientes, estudantes, cidadãos, fornecedores ou parceiros de negócios. Todos em uma organização devem ser um administrador responsável pelos dados sobre as partes interessadas.
- **Regulamentos governamentais**  
Existem regulamentos governamentais para proteger os interesses de algumas partes interessadas. Os regulamentos têm objetivos diferentes. Alguns restringem o acesso às informações, enquanto outros garantem abertura, transparência e responsabilidade. Os regulamentos diferem entre os países, o que significa que as organizações que realizam negócios internacionalmente precisam estar cientes e capazes de atender aos requisitos de proteção de dados onde fazem negócios.
- **Preocupações comerciais proprietárias**  
Cada organização possui dados proprietários para proteger. Os dados de uma organização fornecem informações sobre seus clientes e, quando alavancados com eficiência, podem fornecer uma vantagem competitiva. Se dados confidenciais forem roubados ou violados, uma organização poderá perder vantagem competitiva.
- **Necessidades de acesso legítimo**  
Ao proteger dados, as organizações também devem habilitar o acesso legítimo. Os processos de negócios exigem que indivíduos em determinadas funções possam acessar, usar e manter dados.
- **Obrigações contratuais**  
Acordos contratuais e de não divulgação também influenciam os requisitos de segurança de dados. Por exemplo, o PCI Standard, um contrato entre empresas de cartão de crédito e empresas individuais, exige que certos tipos de dados sejam protegidos de maneiras definidas (por exemplo, criptografia obrigatória para senhas de clientes).



Políticas e procedimentos eficazes de segurança de dados permitem que as pessoas certas usem e atualizem os dados da maneira correta e restrinjam todo acesso e atualizações inadequados.



Fontes de requisitos de segurança de dados (DMBOK2)

## SAIBA MAIS

\* O princípio Cachinhos Dourados é nomeado por analogia à história infantil Os Três Ursos, na qual uma garotinha chamada Cachinhos Dourados prova três tigelas diferentes de mingau e descobre que . [1] Como a história das crianças é bem conhecida em todas as culturas, o conceito de "a quantidade certa" é facilmente compreendido e aplicado a uma ampla gama de disciplinas, incluindo psicologia do desenvolvimento, biologia, [2] astronomia, economia [3] e Engenharia.

Compreender e cumprir os interesses e necessidades de privacidade e confidencialidade de todas as partes interessadas é do interesse de todas as organizações. Todos os relacionamentos com usuários dependem do uso responsável dos dados.



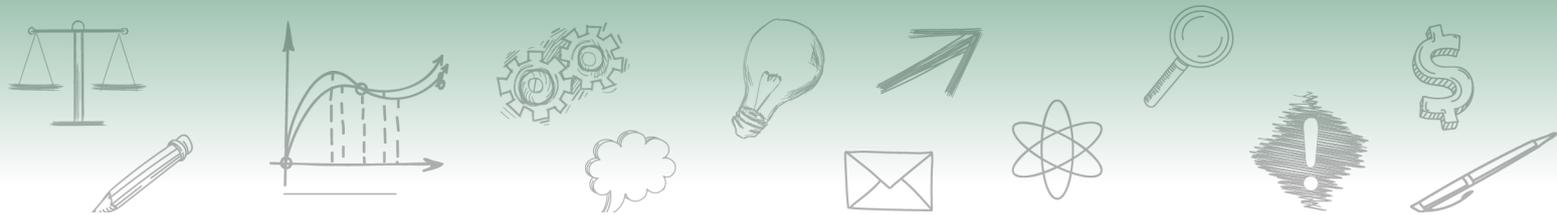
Os objetivos das atividades de segurança de dados incluem:

- Habilitar o acesso apropriado e impedir o acesso inadequado aos ativos de dados corporativos.
- Permitir conformidade com regulamentos e políticas de privacidade, proteção e confidencialidade.
- Garantir que os requisitos das partes interessadas em privacidade e confidencialidade sejam atendidos.

## 5.2 Princípios de segurança de dados

Como os requisitos específicos mudam ao longo do tempo e diferem entre locais, as práticas de segurança de dados devem seguir os princípios orientadores, incluindo:

- **Colaboração**  
A segurança de dados é um esforço colaborativo que envolve gestores de segurança de TI, Gestores de dados / governança de dados, equipes de auditoria interna e externa e o departamento jurídico.
- **Abordagem corporativa**  
Os padrões e políticas de segurança de dados devem ser aplicados de forma consistente em toda a organização.
- **Gerenciamento proativo**  
O sucesso no gerenciamento de segurança de dados deve ser proativo e dinâmico, envolver todas as partes interessadas, gerenciar mudanças e superar gargalos organizacionais ou culturais, como a separação tradicional de responsabilidades entre segurança da informação, tecnologia da informação, gestão de dados e partes interessadas nos negócios.
- **Responsabilidade Clara**  
As funções e responsabilidades devem ser claramente definidas, incluindo a 'cadeia de custódia' dos dados nas organizações e funções.
- **Orientado por metadados**  
A classificação de segurança para elementos de dados é uma parte essencial da definição de dados.
- **Reduza o risco reduzindo a ex...**  
Minimize a proliferação de dados sensíveis/confidenciais, especialmente em ambientes que não sejam de produção.



A redução de riscos e o crescimento dos negócios são os principais impulsionadores das atividades de segurança de dados. Garantir a segurança dos dados de uma organização reduz os riscos e agrega vantagens competitivas. A segurança em si é um ativo de alto valor e há também um imperativo ético para proteger os dados.

Os objetivos de mitigar riscos e aumentar os negócios podem ser complementares e apoiarem-se mutuamente, se integrados a uma estratégia coerente de gerenciamento e proteção de informações.

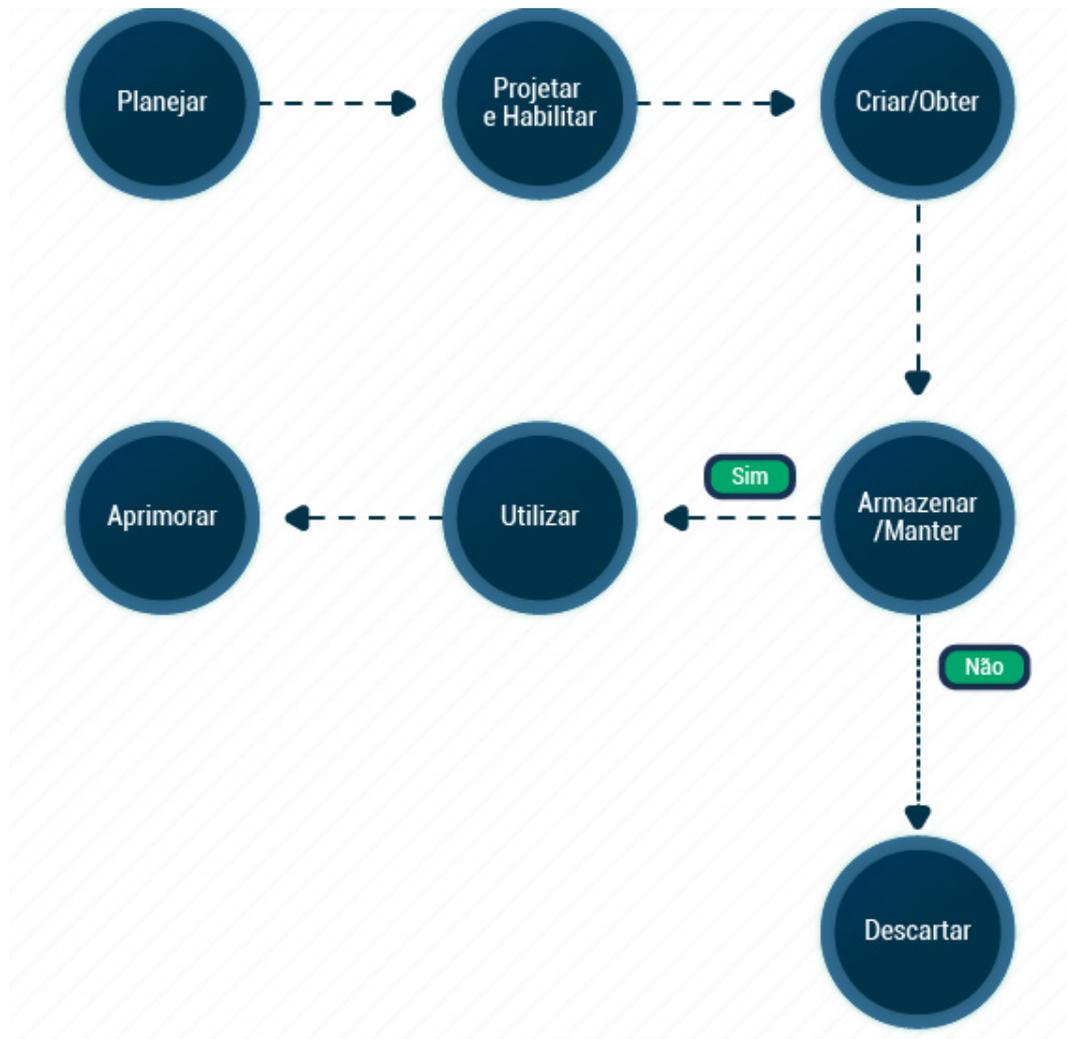
### 5.3. Segurança de dados e gerenciamento de dados corporativos

Na medida em que as regulamentações de dados aumentam, geralmente em resposta a roubos e violações de dados, também aumentam os requisitos de conformidade.

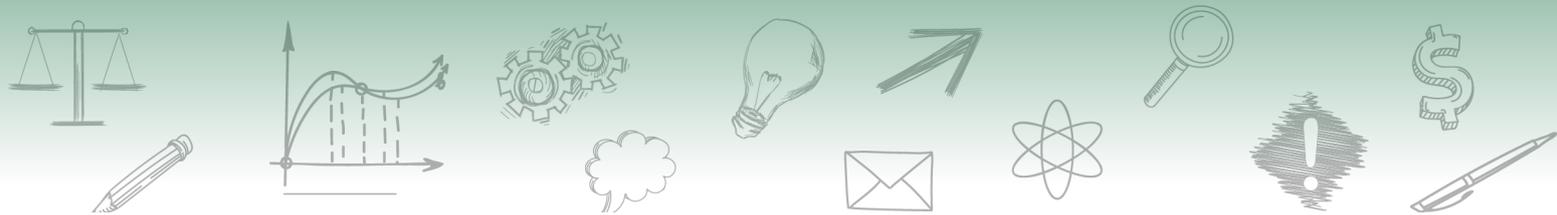
As organizações de segurança geralmente têm a tarefa de gerenciar não apenas os requisitos de conformidade de TI, mas também políticas, práticas, classificações de dados e regras de autorização de acesso em toda a organização.

Como em outros aspectos do gerenciamento de dados, é melhor abordar a segurança dos dados como uma iniciativa corporativa e fazê-lo durante todo o ciclo de vida dos dados. Sem um esforço coordenado, as unidades de negócios encontrarão soluções diferentes para as necessidades de segurança, aumentando o custo geral e potencialmente reduzindo a segurança devido a proteção inconsistente.

Arquitetura ou processos de segurança ineficazes podem custar às organizações devido a violações e perda de produtividade.



- **PLANEJAR:** conectar dados a requisitos de segurança / privacidade.
- **PROJETAR e HABILITAR:** Implemente proteção e segurança de dados nos sistemas
- **CRIAR / OBTER:** Classifique novos dados para que possam ser protegidos adequadamente.
- **ARMAZENAR / MANTER:** Garanta que o armazenamento de dados siga os requisitos da política do governança de dados e da regulação de governo.
- **UTILIZAR:** Gerencie os direitos de acesso para evitar o uso indevido e habilitar o uso apropriado dos dados.
- **APRIMORAR:** Atualize os requisitos relacionados à normas de regulação legal e das ameaças emergentes.



- **DESCARTAR:** Elimine os dados de acordo com os requisitos da políticas de Governança e dos regulamentos pertinentes.

A segurança de dados e de informações começa por avaliar o estado atual dos dados de uma organização para identificar quais dados requerem proteção. O processo inclui as seguintes etapas:

- **Identificar e classificar ativos de dados confidenciais:** Dependendo do setor e da organização, pode haver poucos ou muitos ativos e uma variedade de dados confidenciais - identificação pessoal, saúde, etc.
- **Localizar dados confidenciais em toda a organização:** Os requisitos de segurança podem ser diferentes, dependendo de onde os dados são armazenados. Uma quantidade significativa de dados confidenciais em um único local representa um alto risco devido ao dano possível de uma única violação.
- **Determinar como cada ativo precisa ser protegido:** As medidas necessárias para garantir a segurança podem variar entre ativos, dependendo do conteúdo dos dados e do tipo de tecnologia.
- **Identificar como essas informações interagem com os processos de negócios:** A análise dos processos de negócios é necessária para determinar qual acesso é permitido e sob quais condições.

Além de classificar os dados em si, é necessário avaliar ameaças externas, como as de hackers e criminosos, e os riscos internos apresentados por funcionários e processos. Muitos dados são perdidos ou expostos pela ignorância de quem trabalha diretamente com os dados, que por vezes não percebem que as informações são altamente sensíveis ou ignoram as políticas de segurança.

O impacto de violações de segurança em marcas bem estabelecidas nos últimos anos resultou em enormes prejuízos. Não apenas as ameaças externas da comunidade de hackers criminais estão se tornando mais sofisticadas e direcionadas, mas a quantidade de danos causados por ameaças externas e internas, intencionais ou não, também tem aumentado constantemente ao longo dos anos.

## 5.4 Metadados de Segurança de Dados

Uma abordagem para gerenciar dados confidenciais é via Metadados. Classificações de segurança e sensibilidade regulatória podem ser capturadas ao nível do elemento e do conjunto de dados. Existe tecnologia para marcar dados para que os metadados viajem com as informações à medida que fluem pela organização. O desenvolvimento de um repositório mestre de características de dados significa que todas as partes da organização podem saber exatamente qual nível de informações confidenciais de proteção requer.



Se um padrão comum for aplicado, essa abordagem permitirá que vários departamentos, unidades de negócios e fornecedores usem os mesmos metadados - Segurança padrão. Os metadados podem otimizar a proteção de dados e orientar os processos de uso comercial e suporte técnico, levando a custos reduzidos. Essa camada de segurança pode auxiliar a impedir o acesso não autorizado e o uso indevido de ativos de dados.

Quando dados confidenciais são identificados corretamente como tal, as organizações criam confiança com seus clientes e parceiros. Os próprios metadados relacionados à segurança se tornam um ativo estratégico, aumentando a qualidade das transações, relatórios e análises de negócios, reduzindo o custo da proteção e os riscos associados que as informações perdidas ou roubadas causam.

A classificação de dados é pré-requisito para gerenciar a segurança dos dados. Sua atenção para os dois conceitos que geram restrições de segurança:

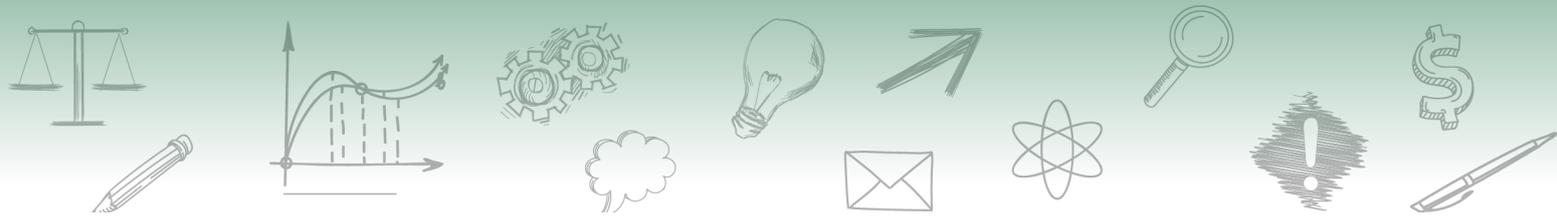
1. **Nível de confidencialidade:** Confidencial significa secreto ou privado. As organizações determinam quais tipos de dados não devem ser conhecidos fora da organização, ou mesmo dentro de certas partes da organização. Informações confidenciais são compartilhadas apenas com base na 'necessidade de conhecimento'. Os níveis de confidencialidade dependem de quem precisa conhecer certos tipos de informações.
2. **Categorias de regulamentação:** são atribuídas com base em normas legais, como leis, tratados, acordos alfandegários e regulamentos. As informações sob regulamentação legal são compartilhadas com base no 'permitido saber'. As maneiras pelas quais os dados podem ser compartilhados são regidos pelos detalhes do regulamento.

A principal diferença entre restrições confidenciais e regulamentares é onde a restrição se origina: as restrições de confidencialidade se originam internamente, enquanto as restrições regulatórias são definidas externamente.

Outra diferença é que qualquer conjunto de dados, como um documento ou uma exibição de banco de dados, pode ter apenas um nível de confidencialidade. Este nível é estabelecido com base no item mais sensível (e mais alto classificado) no conjunto de dados.

As categorizações regulatórias, no entanto, são aditivas. Um único conjunto de dados pode ter dados restritos com base em várias categorias regulatórias. Para garantir a conformidade regulamentar, aplique todas as ações necessárias para cada categoria, juntamente com os requisitos de confidencialidade.

Quando aplicadas ao direito do usuário (a agregação dos elementos de dados específicos aos quais uma autorização do usuário fornece acesso), todas as políticas de proteção devem ser seguidas, independentemente de terem se originado interna ou externamente.



## SAIBA MAIS

**Decreto presidencial dispõe sobre governança no compartilhamento de dados pelo Poder Público**

O Decreto nº 10.046, de 09 de outubro de 2019, que estabelece normas e diretrizes para o compartilhamento de dados entre os órgãos e entidades da administração pública direta, autárquica e fundacional e os demais Poderes da União, foi recentemente sancionado pelo Presidente da República. Referido Decreto atua em caráter complementar à Lei Geral de Proteção de Dados Pessoais (Lei nº 13.709, de 14 de agosto de 2018 – “LGPD”), definindo alguns atributos a serem considerados como dados pessoais compartilhados e estabelecendo os níveis de compartilhamento de dados pessoais no âmbito da administração pública federal, a partir do grau de confidencialidade dos dados que estão sob o controle do Poder Público. Com base nessas definições, serão estabelecidas regras específicas de compartilhamento e medidas de segurança para tais transferências.

O Decreto também institui o Cadastro Base do Cidadão e o Comitê Central de Governança de Dados. O primeiro tem a finalidade de criar uma base de dados integradora única, em formato interoperável, capaz de disponibilizar diversas informações pessoais dos cidadãos aos órgãos e entidades do Poder Executivo Federal; e o segundo, órgão deliberativo, tem como principais competências o estabelecimento de diretrizes e regras para as diferentes categorizações de compartilhamento (amplo, restrito e específico); a avaliação da compatibilidade de políticas de segurança da informação com os compartilhamentos realizados; a avaliação da qualidade dos dados cadastrados; a resolução de controvérsias sobre a validade de informações cadastrais e as regras de prevalência entre eventuais registros administrativos conflitantes, no caso de cruzamento de informações entre bases de dados do Cadastro Base; dentre outras.

Segundo consta no Art. 3º:

(...)

**Art. 3º** O compartilhamento de dados pelos órgãos e entidades de que trata o art. 1º observará as seguintes diretrizes:

**I** - a informação do Estado será compartilhada da forma mais ampla possível, observadas as restrições legais, os requisitos de segurança da informação e comunicações e o disposto na Lei nº 13.709, de 14 de agosto de 2018 - Lei Geral de Proteção de Dados Pessoais;



**II - o compartilhamento de dados sujeitos a sigilo implica a assunção, pelo receptor de dados, dos deveres de sigilo e auditabilidade impostos ao custodiante dos dados;**

**III - os mecanismos de compartilhamento, interoperabilidade e auditabilidade devem ser desenvolvidos de forma a atender às necessidades de negócio dos órgãos e entidades de que trata o art. 1º, para facilitar a execução de políticas públicas orientadas por dados;**

**IV - os órgãos e entidades de que trata o art. 1º colaborarão para a redução dos custos de acesso a dados no âmbito da administração pública, inclusive, mediante o reaproveitamento de recursos de infraestrutura por múltiplos órgãos e entidades;**

**V - nas hipóteses em que se configure tratamento de dados pessoais, serão observados o direito à preservação da intimidade e da privacidade da pessoa natural, a proteção dos dados e as normas e os procedimentos previstos na legislação; e**

**VI - a coleta, o tratamento e o compartilhamento de dados por cada órgão serão realizados nos termos do disposto no art. 23 da Lei nº 13.709, de 2018.**

(...)

**Além disso, no Art. 4º é definido os níveis de compartilhamento:**

(...)

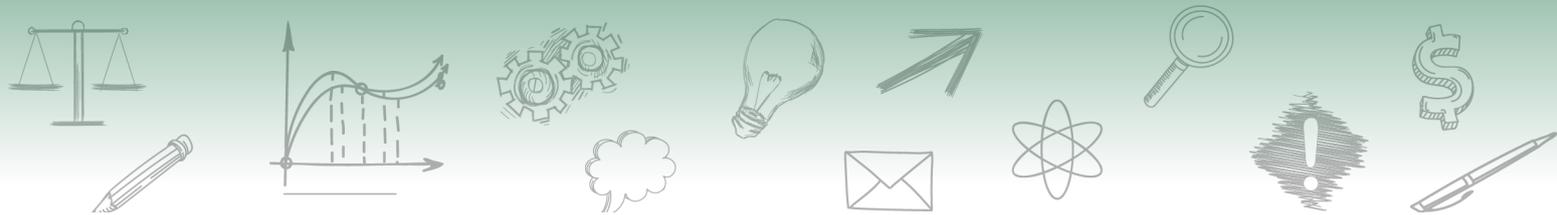
#### **DOS NÍVEIS DE COMPARTILHAMENTO DE DADOS**

**Art. 4º O compartilhamento de dados entre os órgãos e as entidades de que trata o art. 1º é categorizado em três níveis, de acordo com sua confidencialidade:**

**I - compartilhamento amplo, quando se tratar de dados públicos que não estão sujeitos a nenhuma restrição de acesso, cuja divulgação deve ser pública e garantida a qualquer interessado, na forma da legislação;**

**II - compartilhamento restrito, quando se tratar de dados protegidos por sigilo, nos termos da legislação, com concessão de acesso a todos os órgãos e entidades de que trata o art. 1º para a execução de políticas públicas, cujo mecanismo de compartilhamento e regras sejam simplificados e estabelecidos pelo Comitê Central de Governança de Dados; e**

**III - compartilhamento específico, quando se tratar de dados protegidos por sigilo, nos termos da legislação, com concessão de acesso a órgãos e entidades**



específicos, nas hipóteses e para os fins previstos em lei, cujo compartilhamento e regras sejam definidos pelo gestor de dados.

§ 1º A categorização do nível de compartilhamento será feita pelo gestor de dados, com base na legislação.

§ 2º A categorização do nível de compartilhamento será detalhada de forma a tornar clara a situação de cada item de informação.

§ 3º A categorização do nível de compartilhamento como restrito ou específico será publicada pelo respectivo gestor de dados no prazo de noventa dias, contado da data de publicação das regras de compartilhamento de que trata o art. 31.

§ 4º A categorização do nível de compartilhamento como restrito e específico especificará o conjunto de bases de dados por ele administrado com restrições de acesso e as respectivas motivações.

§ 5º A categorização do nível de compartilhamento, na hipótese de ainda não ter sido feita, será realizada pelo gestor de dados quando responder a solicitação de permissão de acesso ao dado.

§ 6º A categorização do nível de compartilhamento será revista a cada cinco anos, contados da data de publicação deste Decreto ou sempre que identificadas alterações nas diretrizes que ensejaram a sua categorização.

§ 7º Os órgãos e entidades de que trata o art. 1º priorizarão a categoria de compartilhamento de dados de maior abertura, em compatibilidade com as diretrizes de acesso a informação previstas na legislação.

(...)

As regras gerais de compartilhamento encontram-se detalhadas a partir do Artigo 5º do Decreto nº 10.046, 09/10/2019: [http://www.planalto.gov.br/ccivil\\_03/\\_Ato2019-2022/2019/Decreto/D10046.htm](http://www.planalto.gov.br/ccivil_03/_Ato2019-2022/2019/Decreto/D10046.htm).

## 5.5 Arquitetura de segurança de dados

A arquitetura corporativa define os ativos e componentes de informações de uma empresa, suas inter-relações e regras de negócios relacionadas à transformação, princípios e diretrizes.

A arquitetura Data Security é o componente da arquitetura corporativa que descreve como a segurança dos dados é implementada dentro da organização para satisfazer as regras de negócios e os regulamentos externos.



Atenção, a arquitetura influencia:

- Ferramentas usadas para gerenciar a segurança dos dados.
- Padrões e mecanismos de criptografia de dados.
- Diretrizes de acesso externos.
- Protocolos de transmissão de dados.
- Documentos necessários.
- Padrões de acesso remoto.
- Procedimentos de comunicação de incidentes de violação de segurança.

A arquitetura de segurança é particularmente importante para a integração de dados entre:

- Sistemas internos e unidades de negócios.
- Uma organização e seus parceiros de negócios externos.
- Uma organização e agências reguladoras.

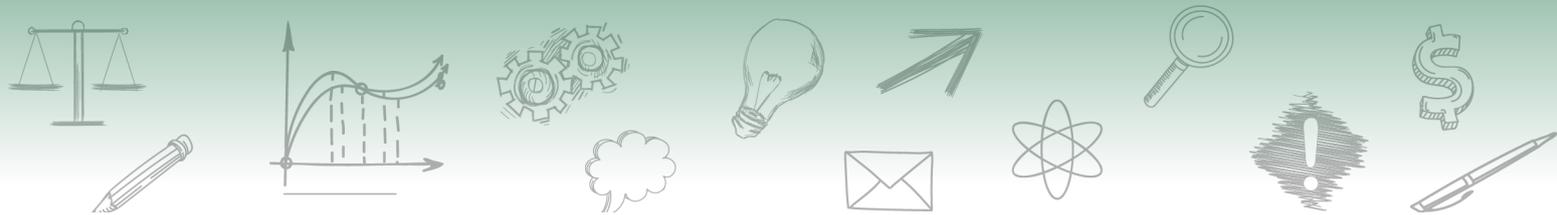
Por exemplo, um padrão de arquitetura referente a um mecanismo de integração orientado a serviços entre partes internas e externas exigiria uma implementação de segurança de dados diferente da arquitetura de integração tradicional de intercâmbio eletrônico de dados (*Electronic Data Interchange* - EDI).

Para uma organização de grande porte, a função formal de ligação entre essas disciplinas é essencial para proteger as informações contra uso indevido, roubo, exposição e perda. Cada parte deve estar ciente dos elementos que interessam às outras, para que possam falar uma linguagem comum e trabalhar em direção a objetivos compartilhados.

## 5.6 Planejando a segurança de dados

O planejamento da segurança inclui o planejamento do processo, bem como a classificação dos dados e o planejamento da arquitetura. Inclui segurança não apenas de sistemas, mas também de instalações, dispositivos e credenciais.

A implementação de boas práticas começa com a identificação de requisitos. Eles se baseiam amplamente em regulamentos para indústrias e regiões geográficas específicas. É importante garantir que uma organização possa atender a requisitos que possam ser direcionados por aqueles com quem interage; por exemplo, a União Europeia possui requisitos de privacidade mais rigorosos do que os Estados Unidos. Os requisitos também serão baseados nos riscos relacionados ao cenário do sistema da própria organização.



Os requisitos devem ser formalizados em políticas e apoiados por padrões coesos como níveis de classificação. Políticas e padrões precisam ser mantidos e aperfeiçoados à medida em que os regulamentos evoluem. Treinamento contínuo, acesso a dados e o uso do sistema precisarão ser monitorados para garantir a conformidade.

A cultura corporativa influencia profundamente a maneira como mantemos os dados seguros. As organizações geralmente acabam reagindo a crises, em vez de gerenciar proativamente a responsabilidade e garantir a auditoria. Embora a segurança perfeita dos dados seja quase impossível, a melhor maneira de evitar violações da segurança dos dados é conscientizar e entender os requisitos, políticas e procedimentos de segurança.

As organizações podem aumentar a conformidade, por meio de:

- **Treinamento**  
Promoção de padrões por meio de treinamento em iniciativas de segurança em todos os níveis da organização. Siga o treinamento com mecanismos de avaliação, como testes on-line focados em melhorar a conscientização dos funcionários. Esse treinamento e teste devem ser obrigatórios e um pré-requisito para a avaliação de desempenho dos funcionários.
- **Políticas consistentes**  
Definição de políticas de segurança de dados e políticas de conformidade regulamentar para grupos de trabalho e departamentos que complementam e se alinham às políticas da organização. A adoção de uma mentalidade "agir localmente" ajuda a envolver as pessoas de maneira mais ativa.
- **Avalie os benefícios de segurança**  
Vincule os benefícios da segurança de dados a iniciativas organizacionais. As organizações devem incluir métricas objetivas para as atividades de segurança de dados em suas métricas de *balanced scorecard* e avaliações de projetos.
- **Definir requisitos de segurança para fornecedores**  
Inclua requisitos de segurança de dados em contratos de nível de serviço e obrigações contratuais de terceirização. Os contratos de *Service Level Agreement (SLA)* devem incluir todas as ações de proteção de dados.
- **Crie um senso de urgência**  
Enfatize os requisitos legais, contratuais e regulamentares para criar um senso de urgência e uma estrutura interna para o gerenciamento de segurança de dados.
- **Comunicações contínuas**  
Suporte a um programa contínuo de treinamento de segurança dos funcionários, informando-os sobre práticas seguras de computação e ameaças atuais. Um programa em andamento comunica que a computação segura é importante o suficiente para ser suportada pela gerência.



## Referências

Navigating the Labyrinth: An Executive Guide to Data Management by Laura Sebastian-Coleman for DAMA International Publisher: Technics Publications Release Date: June 2018.

DAMA-DMBOK: Data Management Body of Knowledge (2nd Edition) by DAMA International Publisher: Technics Publications, Release Date: July 2017.