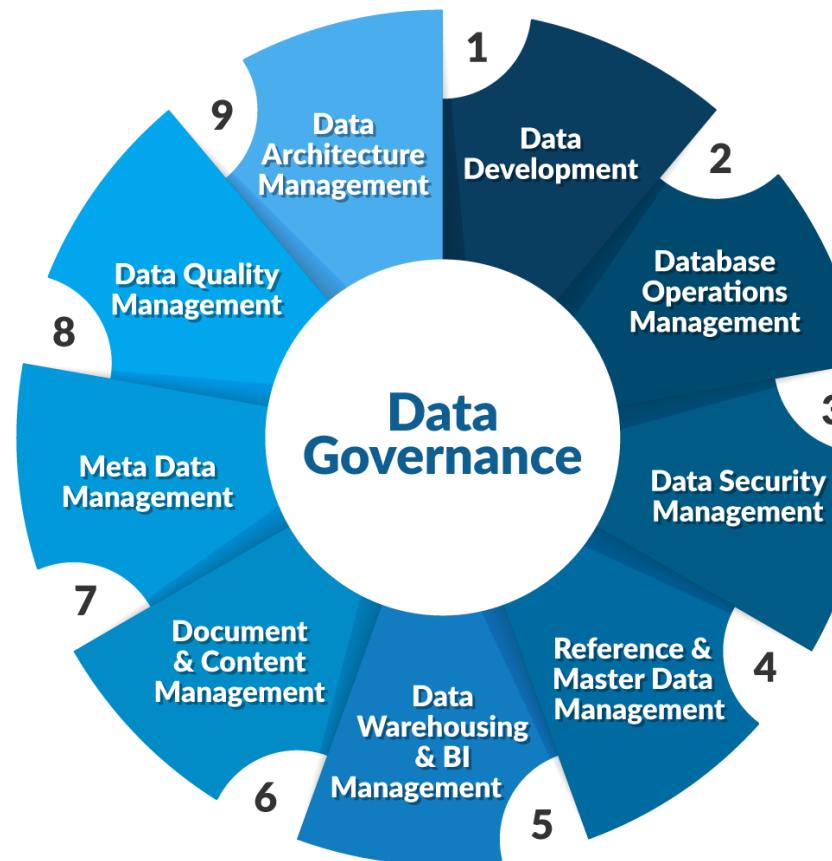


GOVERNANÇA DE DADOS



PARTE I- DEFINIÇÃO DE GOVERNANÇA DE DADOS, IMPORTÂNCIA E FRAMEWORK



DEFINIÇÃO DE GOVERNANÇA DE DADOS

Data Governance, Data Stewardship, Data Quality.

- A Governação de Dados é uma coleção de práticas e processos que ajudam a garantir o gerenciamento formal de ativos de dados dentro de uma organização.
- Inclui métodos, tecnologias e comportamentos relacionados ao gerenciamento adequado dos dados.

Também lida com:

- Segurança
- Privacidade
- Integridade
- Usabilidade
- Integração
- Conformidade
- Disponibilidade
- Funções
- Responsabilidades
- Gerenciamento geral dos fluxos de dados internos e externos da organização.

IMPORTÂNCIA

✓ Qualidade dos dados

✓ Colaboração

✓ Eficiência de Análise

A importância também está relacionada com:

- Redução de custos
- Procedimentos mais precisos
- Maior transparência
- Melhor treinamento e práticas educacionais
- Aumento do valor
- Sistemas de dados padronizados
- Resolução de problemas passados e atuais
- Monitoramento e rastreamento aprimorados
- Crescimento da receita da empresa

Os 3 Componentes Principais da
Governança de Dados

✓ Backup (cópias de segurança)

✓ Archiving (seleção dos tipos de dados que serão
guardados)

✓ E-Discovery (saber o que está disponível)

A área de negócio deve ser a responsável pela
Governança de Dados, visto que a mesma é um processo
de estratégia corporativa!

Cabe a área de TI fazer o gerenciamento!

Resultados da Governança de Dados:

- Definições de qualidade dos dados
- Glossário de negócios
- Funções e responsabilidades
- Catálogos de dados governados
- Criação de metadados

FRAMEWORK

- Data Ownership – Policies, Guidelines, & Standards
- Data Stewardship – Data Quality Framework
- Role Definition and Accessibility – Privacy, Compliance, & Security
- Reliable Flow of Information – Information Architecture & Integrations
- Knowledge from information – Reporting & Analytics

Data Driven Decisions
and
Measurable Outcomes

People, Processes and Technology

5 FASES DE MATURIDADE EM GOVERNANÇA DE DADOS

1 – Chaotic

- No standards
- Reactive approach
- No master data plan
- No strategy

2 – Reactive

- Standards established
- Basic DQM Process established
- Master data plan identified
- Strategy defined and communicated

3 – Defined

- KPIs identified & measured
- Data dictionary and rule dictionary documented and maintained
- N-Tiered stewardship established
- Master data plan executed
- Supporting technology framework deployed
- Root cause for issues being tracked and measured

4 – Proactive

- Continuous improvement feedback loops operating
- Root cause analysis feeding into feedback process
- Proactive approach to management os data and rules dictionary
- DQM process automating measurement of function performance
- All information silos fully integrated with master data systems

5 – Predictive

- Process feedback loops are tuning as opposed to fixing
- DQM processes fully automated with complete audit trail
- Top-down strategy fully in tune with the bottom up application of stewardship => complete cultural alignment across the enterprise
- People, Process, and Techonogy operating in harmony

ESTRATÉGIA DE DADOS



DAMA – International / Brasil
The Global Data Management Community
www.dama.org

PARTE II- ESTRATÉGIA DE IMPLEMENTAÇÃO DE GOVERNAÇA DE DADOS

1. Definição da Estratégia
2. Escolha do Modelo de Governança
3. Seleção do Comitê de Governança
4. Criação do Data Governance Office
5. Assessment (Avaliação do Estado Atual)
6. Definição de Políticas e Procedimentos
7. Design Funcional
8. Processo de Melhoria Contínua

1 . DEFINIÇÃO DA ESTRATÉGIA

- Escrever uma carta de governança de dados (**Governance Charter**) com apoio das partes interessadas e dos envolvidos na implementação da Governança de Dados.
- Essa carta deve conter uma declaração de missão, explicação dos objetivos gerais e deve ser assinada para implementação oficial. Deverá ser incluído quem tem autoridade para fazer o que com os dados e como será a abrangência da Governança. (Template)

2. ESCOLHA DO MODELO DE GOVERNANÇA

Um modelo comum que adota uma abordagem de três níveis inclui:

- **Comitê de Governança** – Grupo de executivos de nível senior para as decisões de alto nível.
- **Data Governance Office** – Grupo de gerenciamento intermediário, que oferece orientação.
- **Data Governance Working Group** – Grupo de trabalho administrativo que contém especialistas no assunto, tanto em negócios quanto em tecnologia da Informação (TI).

5 - ASSESSMENT (AVALIAÇÃO DO ESTADO ATUAL)

- Um Assessment é uma atividade realizada pela **Data Governance Office** para avaliar a situação atual da utilização dos dados e a partir daí traçar planos de ação, políticas, procedimentos e regras para a Governação de Dados.
- Existem 4 tipos principais de assessments:
 - **Cultural:** avalia a maturidade no uso dos dados
 - **Capacidade:** avalia os recursos disponíveis para trabalhar com os dados
 - **Colaborativo:** avalia o nível de colaboração da empresa
 - **Processos:** avalia o nível dos processos de gestão

LEVAMENTOS QUE DEVEM SER FEITOS DURANTE O ASSESSMENT

1. Já existe iniciativa MDM (Master Data Management)?
2. Já existe iniciativa de gestão de Documentos (Content Document Management)?
3. Qual o nível de qualidade dos dados?
4. A empresa já pratica EIM (Enterprise Information Management)?

6 - DEFINIÇÃO DE POLÍTICAS E PROCEDIMENTOS

- Melhorar a qualidade dos dados inclui instituir melhores práticas de gerenciamento de dados através de políticas, procedimentos e padrões, isso deve ocorrer com patrocínio do nível executivo.
- As políticas estarão em um documento dinâmico que será atualizado à medida que as estratégias e metas da empresa mudam.

7 - DESIGN FUNCIONAL

- Analisa funções e processos (não as pessoas) que necessários para implementar o programa de controle de dados.
- Essa equipe irá desenvolver uma **lista** do que a Governança de dados realizará, como **funções e processos de gerenciamento e responsabilidades**, e certificar-se de que está sendo disseminada para todos os envolvidos.
- Objetivo:
 - Determinar os princípios básicos da informação
 - Determinar um baseline das políticas de Governança de Dados
 - Identificar e refinar processos
 - Definir novos processos relacionados a dados
 - Estabelecer critérios e métricas de qualidade dos dados (**o que não pode ser medido não pode ser gerenciado**)

8 - PROCESSO DE MELHORIA CONTÍNUA

- Os processos, as pessoas e a tecnologia, são implantados para maximizar a disponibilidade e integridade dos dados, no entanto esse processo é um ciclo e sempre há espaço para melhoria.
- Deve haver um mecanismo de feedback no processo.
- Representar graficamente os sucessos e as falhas no processo (medir).
- O sucesso da Governança de dados depende das pessoas, se elas sabem seus papéis, responsabilidades e regras, tem foco nos dados e são suportados pela tecnologia que facilita seus trabalhos, a Governança de Dados mostra todo seu valor!

PARTE III – TECNOLOGIAS, FERRAMENTAS E LEI DE PROTEÇÃO AOS DADOS E GDPR

- Os 5 pilares da Governação de dados:
 1. Segurança
 2. MDM (Master Data Management)
 3. Data Quality
 4. Metadados e Glossários
 5. Data Lineage

2 . M D M (M A S T E R D A T A M A N A G E M E N T)

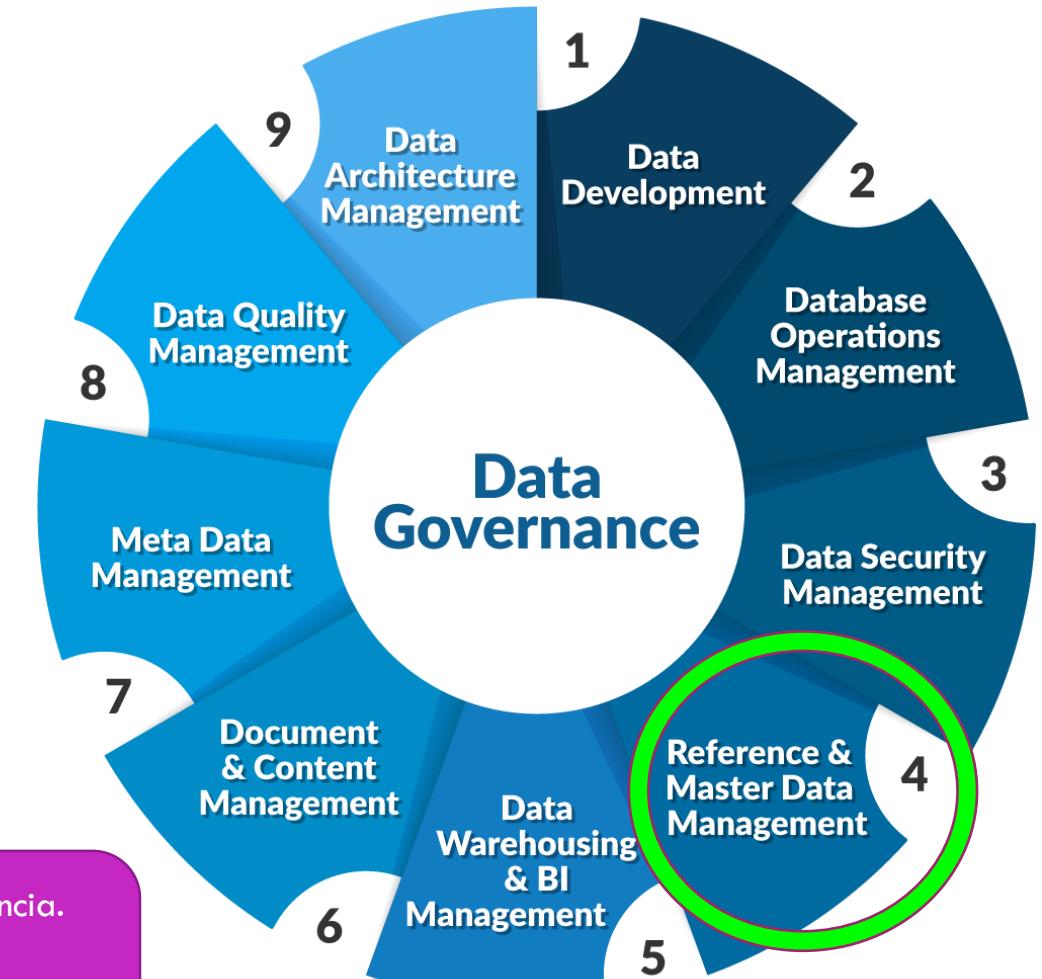
- O MDM é a tecnologia, as ferramentas e os processos que garantem que os dados principais sejam coordenados em toda a empresa. É uma forma de uma organização padronizar e garantir a qualidade dos dados, criando uma **única fonte de referência mestre** para todos os dados críticos, levando a menos erros e menos redundância nos processos de negócio.

Criar a Lista Mestre

- Limpar e padronizar os dados
- Correspondência de dados de todas as fontes para consolidar duplicatas.

Manter a Lista Mestre

- Manter uma cópia única de referência.
- Múltiplas cópias da lista mestre.
- Em ambos os casos anteriores , criar um processo de “Merge” contínuo.



4 . M E T A D A D O S E G L O S S Á R I O S

- Metadados são dados sobre os dados, são o tipo de informação que informa os usuários sobre os dados em si. Alguns elementos típicos:
 1. Título e descrição
 2. Tags(etiquetas) e categorias
 3. Quem criou e quando
 4. Quem modificou pela última vez
 5. Quem pode acessar ou atualizar

PRINCIPAIS PLATAFORMAS USADAS EM GOVERNANÇA DE DADOS

- Alation Data Catalog 4.x
- Collibra Data Governance Center 5.3
- DATUM Information Value Management
- Erwin DG
- IBM Stewardship Ctr & Information Governance Catalog
- Infogix Data3Sixty Govern
- Informatica Data Director
- Informatica Data Governance
- Orchestra Networks EBX5
- SAP Master Data Governance
- Utopia

FERRAMENTAS DE GOVERNANÇA DE DADOS

Lista de ferramentas proprietárias:

- informatica.com (uma série de ferramentas para governança de dados)
- collibra.com
- alation.com
- datumstrategy.com
- orchestranetworks.com (TIBCO EBX)

GERENCIAMENTO DE DOCUMENTOS (DMS) X GERENCIAMENTO DE CONTEÚDO (CMS)

- DMS é um Sistema que ajuda as empresas a criar, rastrear e armazenar documentos digitalizados. Um DMS é usado para classificar, reter e proteger informações eletrônica. Ele também suporta controle de versão, colaboração e fluxos de trabalho. (**DMS gerencia dados estruturados**)
- Sistemas CMS são usados para criar e gerenciar vários tipos de conteúdo digital, além de documentos no sentido tradicional. Por ex, além de ser usado para PDF's, arquivos Word ou arquivos excel, um CMS pode ser usado para gerenciar coisas como imagens, páginas web, registros e arquivos flash. (**CMS gerencia dados estruturados e não estruturados**)

G O V E R N A N Ç A D E D A D O S E B I G D A T A - G E S T Ã O D E M E T A D A D O S

- Big Data pode ser definido pelos seus 4V's: Volume, Variedade, Velocidade e Veracidade.
- Aspecto técnico mais relevante em Governança de Big Data: METADADOS
- Enquanto os dados estão crescendo e o Data Lake está sendo usado pela empresa, surgem as seguintes perguntas:

Quais dados estão disponíveis no data lake, como é consumido/preparado/transformado, quem está usando esses dados, quem está contribuindo para esses dados, quantos anos tem os dados? etc.

- Uma camada de metadados bem mantida pode efetivamente responder estes tipos de consultas e, assim, melhorar a usabilidade do Data Lake.

BENEFÍCIOS DE CAMADA DE METADADOS NA GOVERNANÇA DO BIG DATA

- A camada de metadados captura informações vitais sobre os dados e como eles entram no Data Lake, indexando as informações para que possamos pesquisar por metadados antes de acessarmos os dados diretamente. Capturar metadados é fundamental para tornar os dados mais acessíveis e extrair valor do Big Data.
- Assim a organização aproveitará o potencial do Data Lake e entregar os seguintes mecanismos para os usuários finais acessarem dados e realizar análise:
 - BI de autoatendimento (SSBI)
 - Data-as-a-Service (DaaS)
 - Aprendizagem de Máquina como Serviço
 - Provisionamento de dados (DP)

PRINCIPAIS PLATAFORMAS USADAS EM GOVERNANCE DE BIG DATA

- Cloudera Navigator: governança de dados com Apache Hadoop
- Apache Atlas (*gratuito): Data Governance and Metadata framework for Hadoop
- Apache Falcon: é uma ferramenta de gestão (não de governança especificamente)
- Waterline Data (paga): + Mapr = Plataforma de governança

APACHE ATLAS: DATA GOVERNANCE AND METADATA FRAMEWORK FOR HADOOP

- Principal ferramenta para Governança de Dados no Ecosistema Hadoop. Possui a gestão de metadados.
- Cloudera + Hortonworks
- **Atlas é um conjunto escalonável e extensível de serviços de governança – permitindo que as empresas atuem e atendam de forma eficiente seus requisitos de conformidade de um Data Lake com Hadoop e permite integração com todo ecossistema de dados.**

PRINCIPAIS CARACTERÍSTICAS E FUNCIONALIDADES DO APACHE ATLAS

Classificação dos dados

- Importar ou definir taxonomia orientada para os negócios como anotações para os dados.
- Definir, anotar e automatizar a captura de relacionamentos entre conjuntos de dados e elementos subjacentes, incluindo processo de origem, destino e derivação.
- Exportar metadados para sistemas de terceiros.

Auditoria centralizada

- Captura informações de acesso de segurança para cada aplicação, processo e interação com dados.
- Captura as informações operacionais para execução, etapas e atividades.

Mecanismo de Segurança e Política

- Racionaliza a política de conformidade em tempo de execução com base em dados, esquemas de classificação, atribuições e papéis.
- Definição avançada de políticas para prevenção de dados e derivação baseada na classificação.
- Mascaramento de nível de coluna e linha com base nos valores da célula e atributos.

Pesquisa e Linhagem

- Caminhos de navegação pré-definidos para explorar os dados de classificação e auditoria.
- Localização de dados relevantes e evento de auditoria no Data Lake com rapidez e precisão.
- Permite aos usuários aplicarem drill-down em informações operacionais, de segurança e procedência relacionadas.

LEI DE PROTEÇÃO AOS DADOS

Europa



Brasil - 2020

