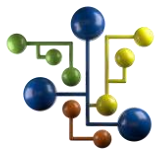


Data Science  
Academy

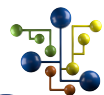
Data Science Academy [jaquelinefernandes6@gmail.com](mailto:jaquelinefernandes6@gmail.com) 5e84edfee32fc3397c122ab7

# Governança de Dados





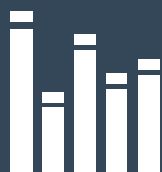
Data Science  
Academy



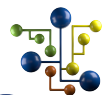
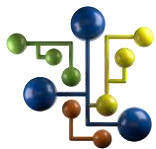
Data Science  
Academy

Data Science Academy jaquelinefernandes6@gmail.com 5e84edfee32fc3397c122ab7

# Os 5 Pilares da Governança de Dados

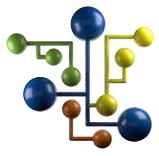


**Governança de Dados**  
O Que é Governança de Dados?

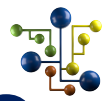


# Os 5 Pilares da Governança de Dados





Data Science  
Academy

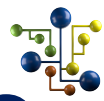
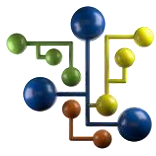


Data Science  
Academy

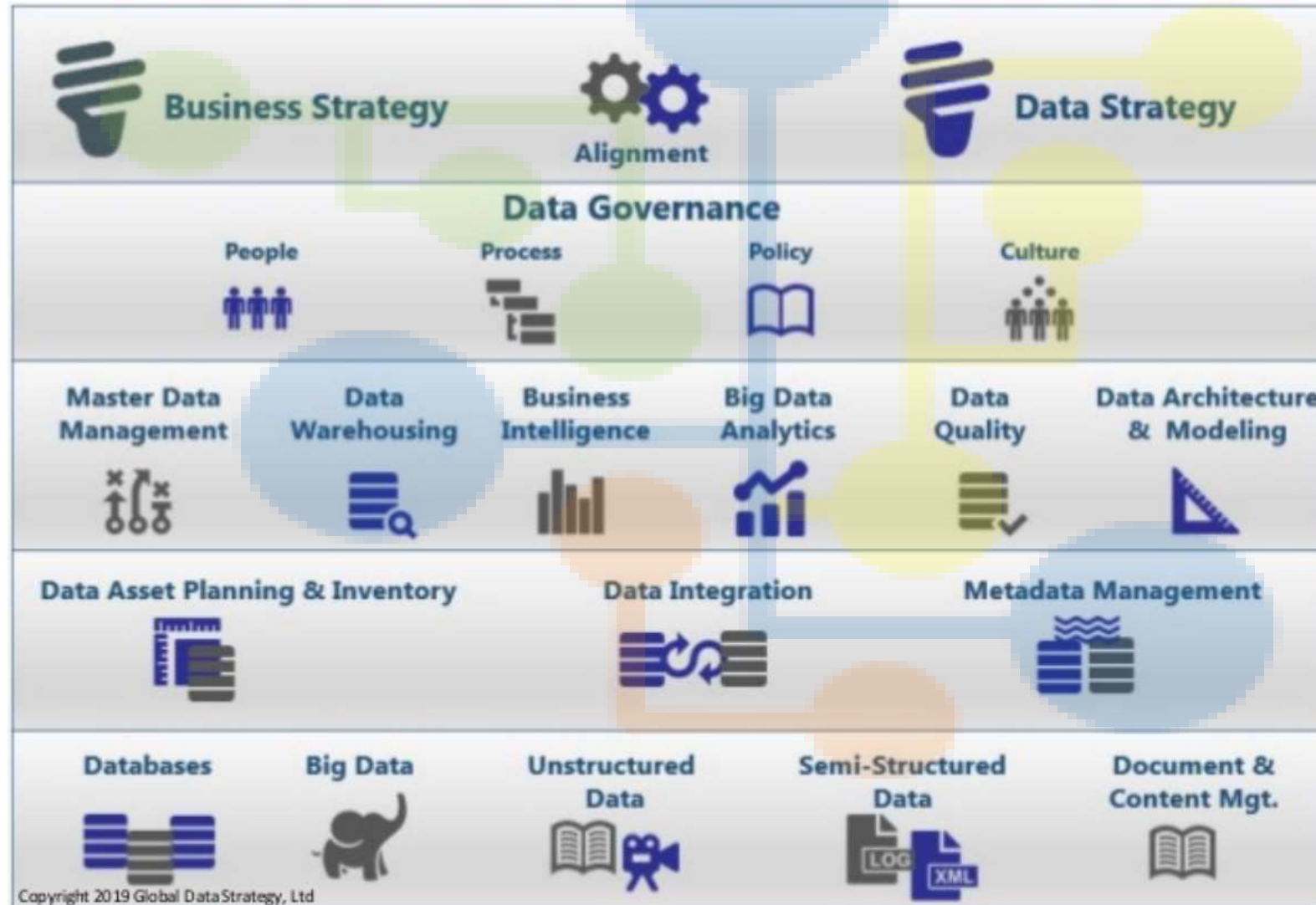
Data Science Academy jaquelinefernandes6@gmail.com 5e84edfee32fc3397c122ab7

# Os 5 Pilares da Governança de Dados

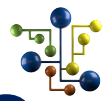
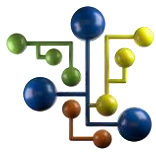




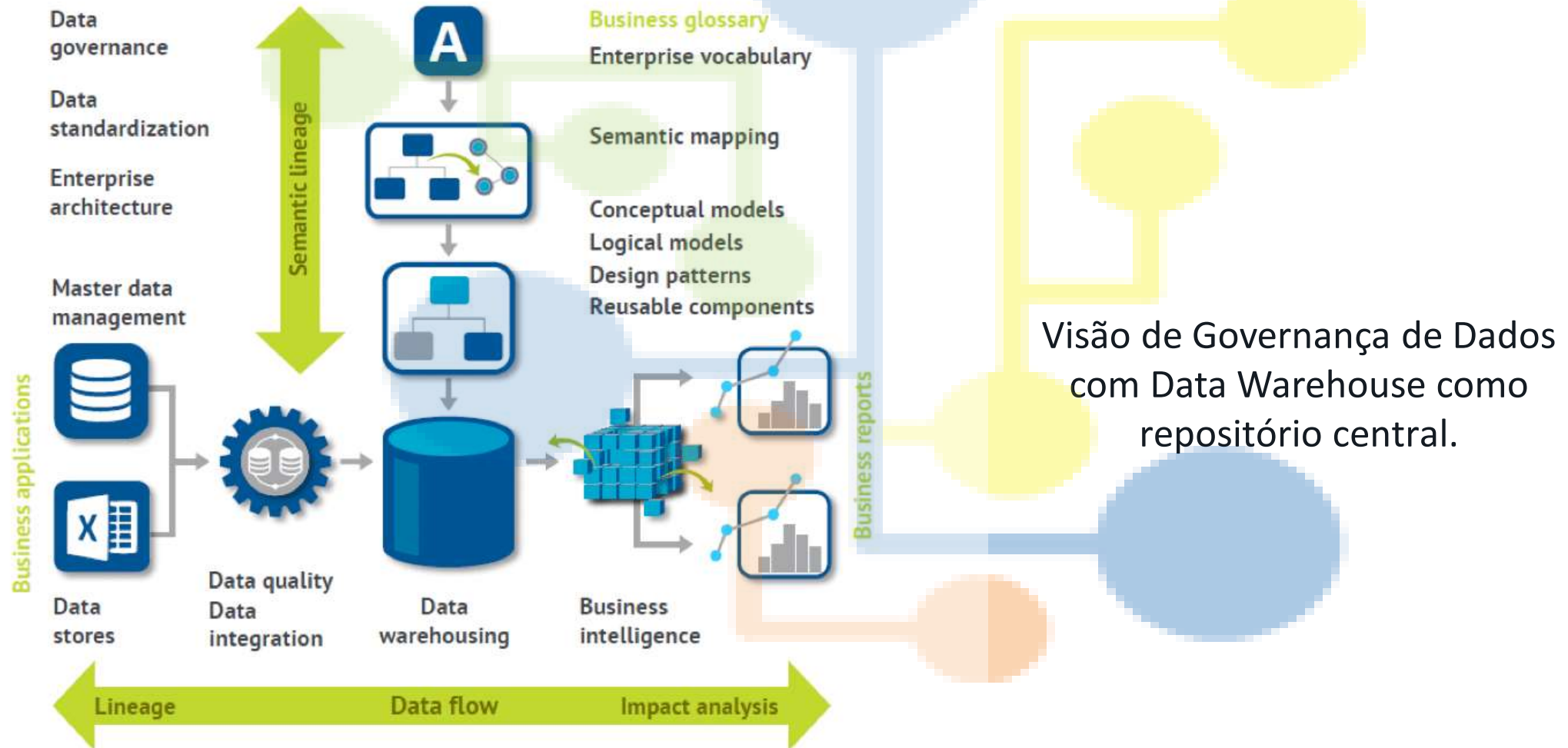
# Os 5 Pilares da Governança de Dados

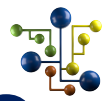
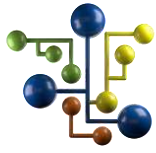




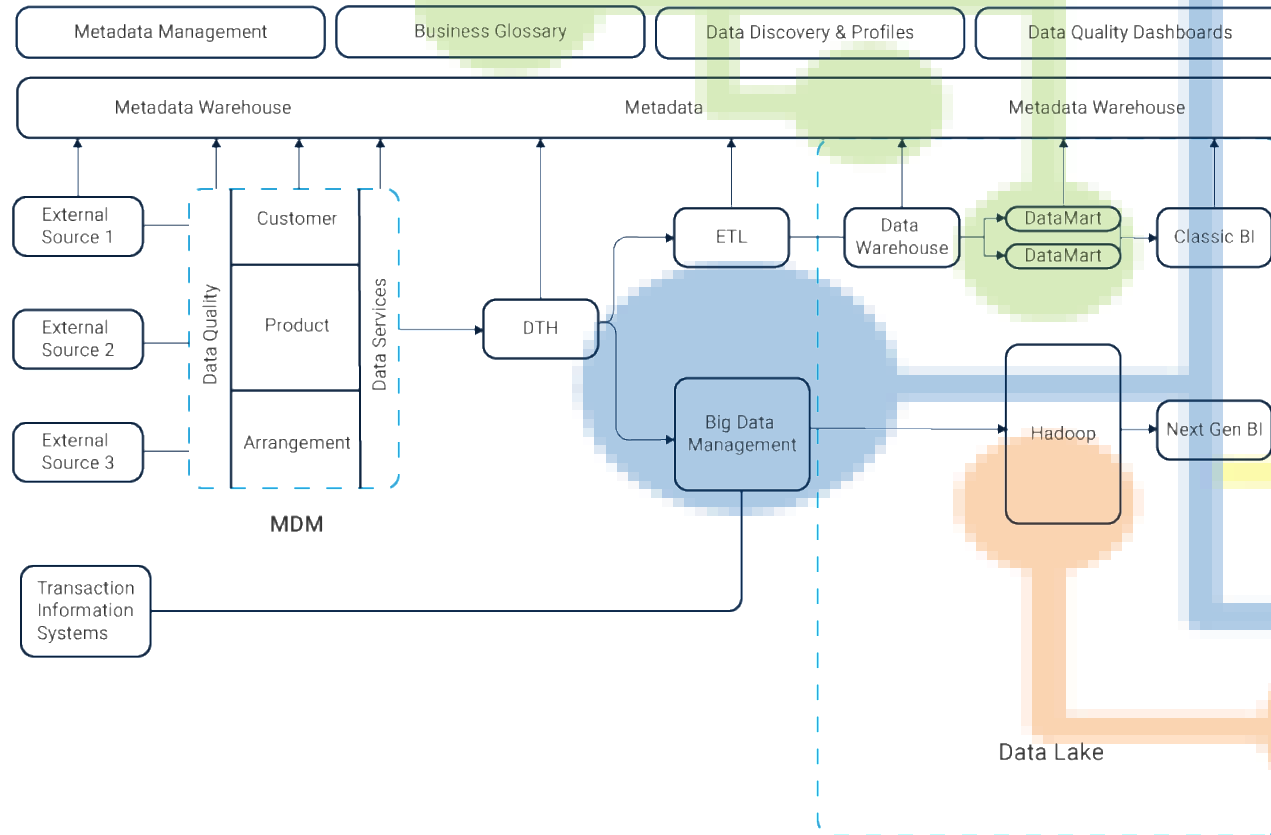


# Os 5 Pilares da Governança de Dados

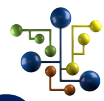
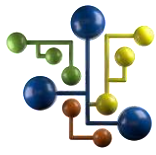




# Os 5 Pilares da Governança de Dados



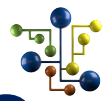
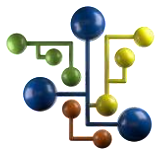
Visão de Governança de Dados  
com Data Lake como repositório  
central.



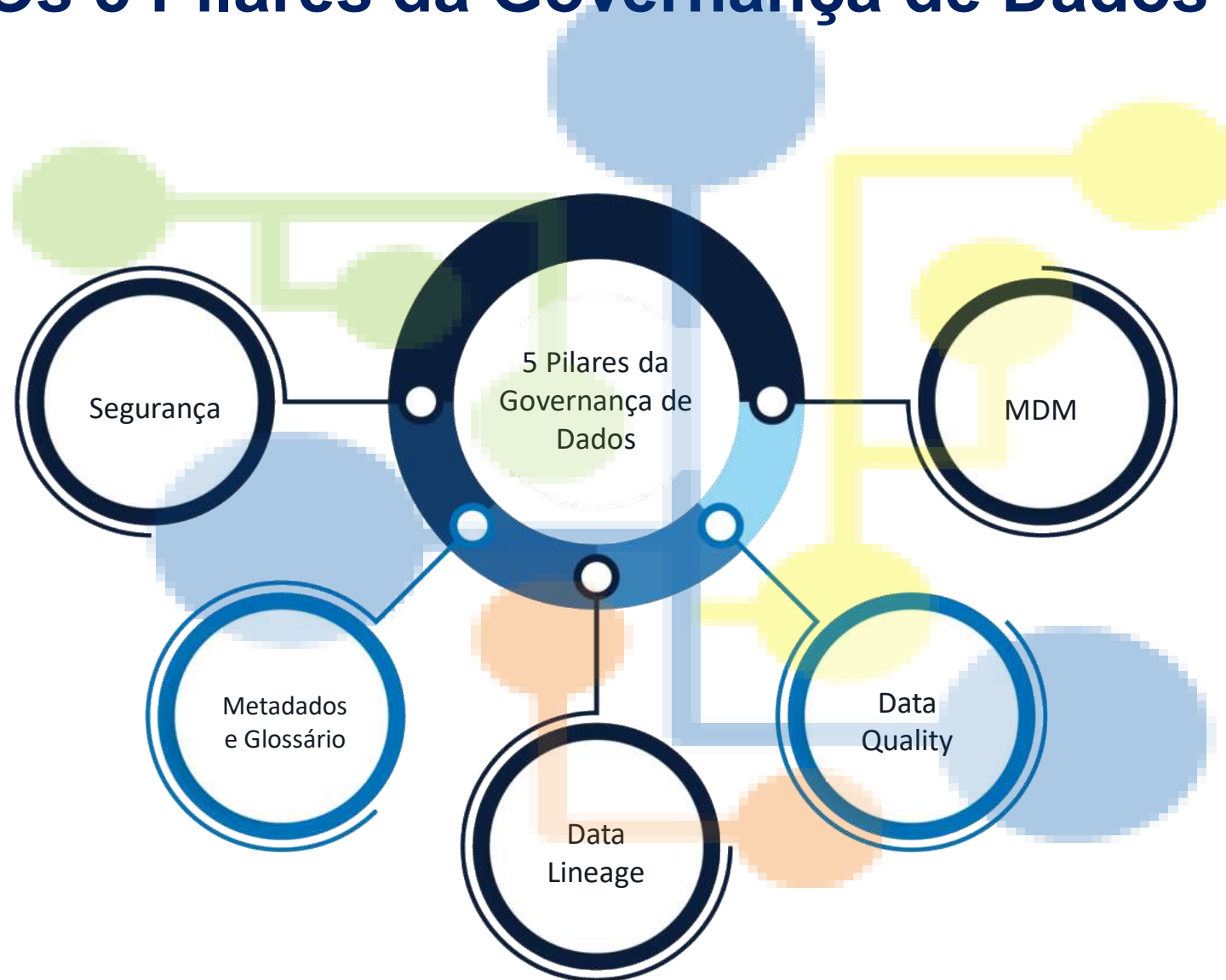
# Os 5 Pilares da Governança de Dados

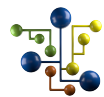
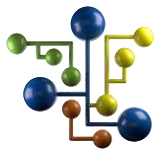




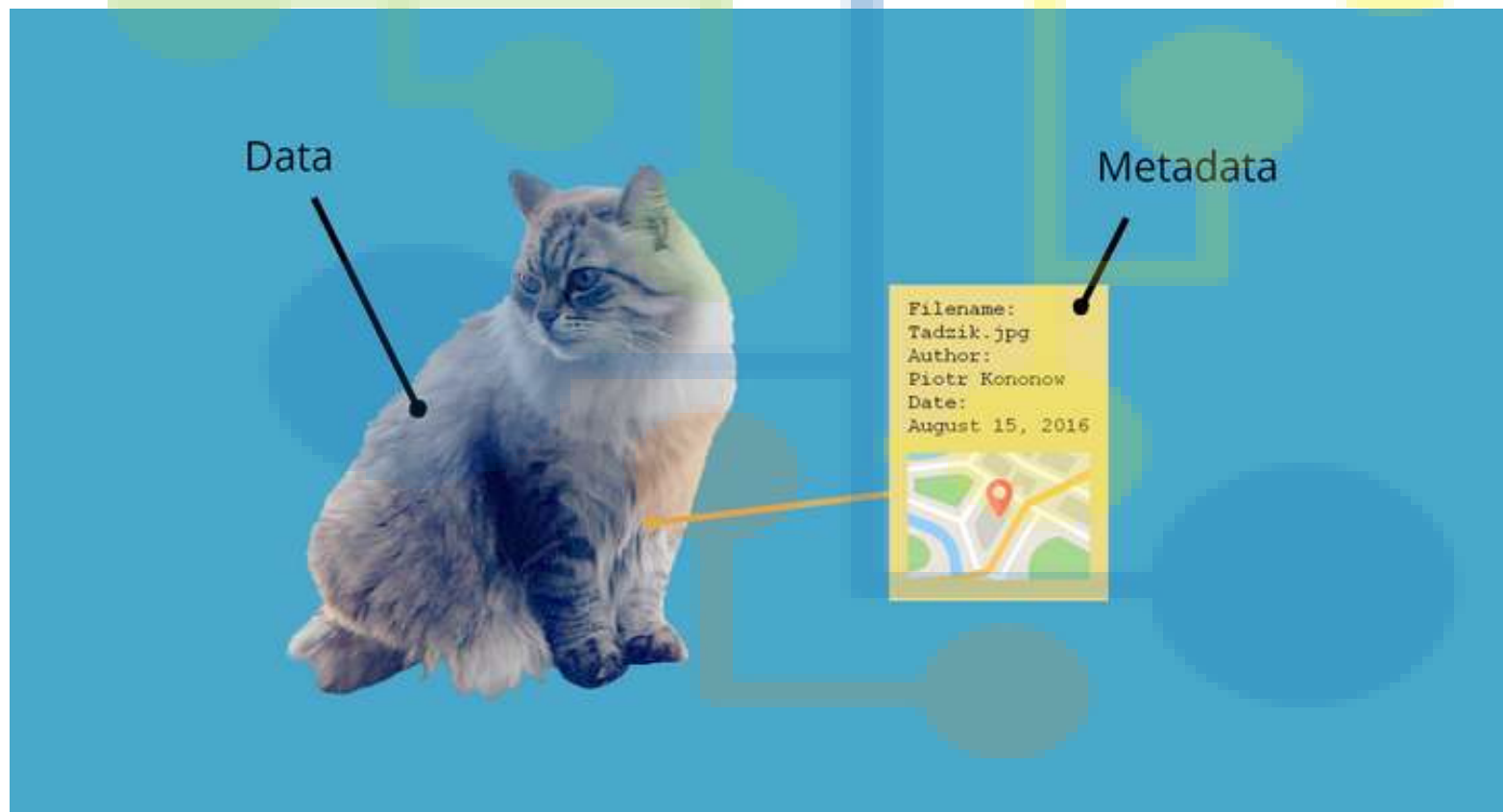


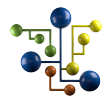
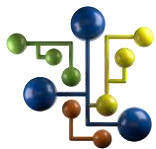
# Os 5 Pilares da Governança de Dados





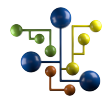
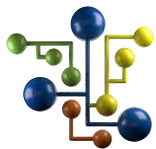
# O Que São Metadados?





# O Que São Metadados?

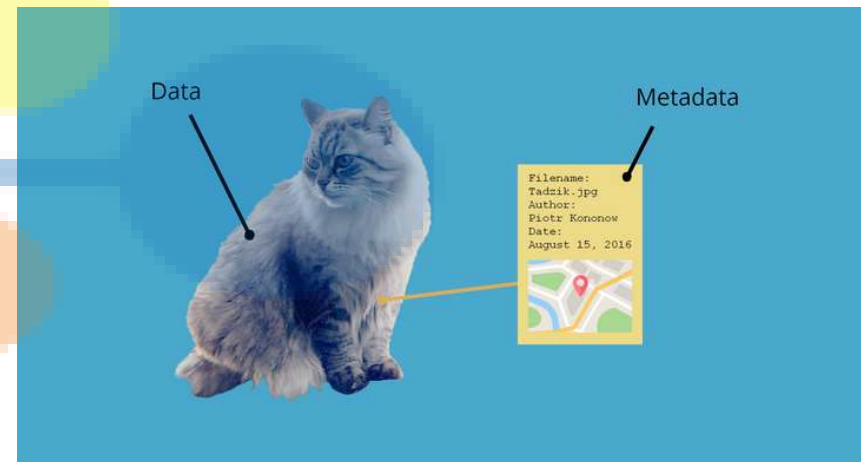
Entre as várias classificações de dados que são vistas nos procedimentos modernos da Ciência de Dados, os metadados são o tipo que informa os usuários sobre os dados em si.

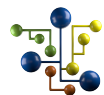
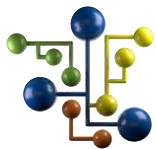


# O Que São Metadados?

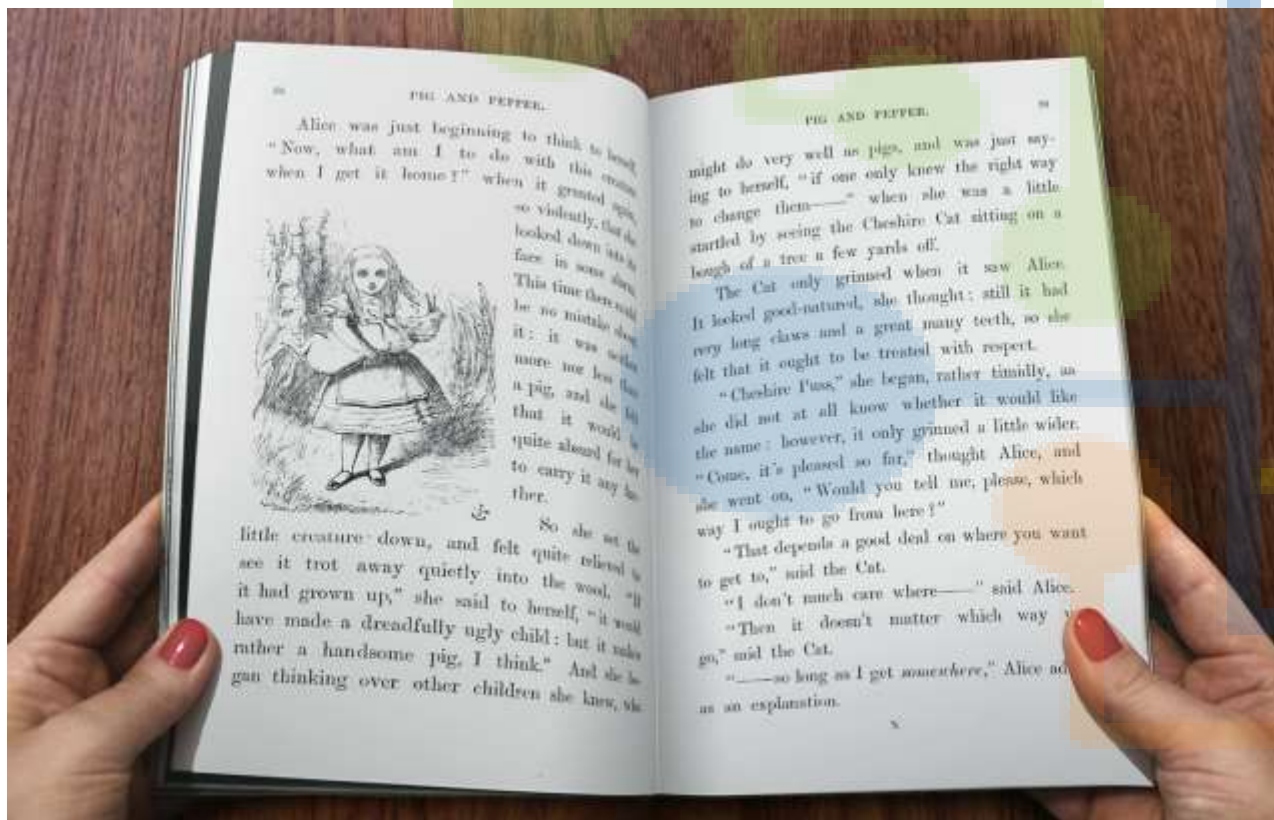
Esses são alguns elementos típicos de metadados:

- Título e descrição
- Tags e categorias
- Quem criou e quando
- Quem modificou pela última vez e quando
- Quem pode acessar ou atualizar



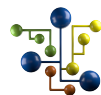
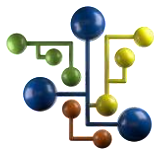


# O Que São Metadados?

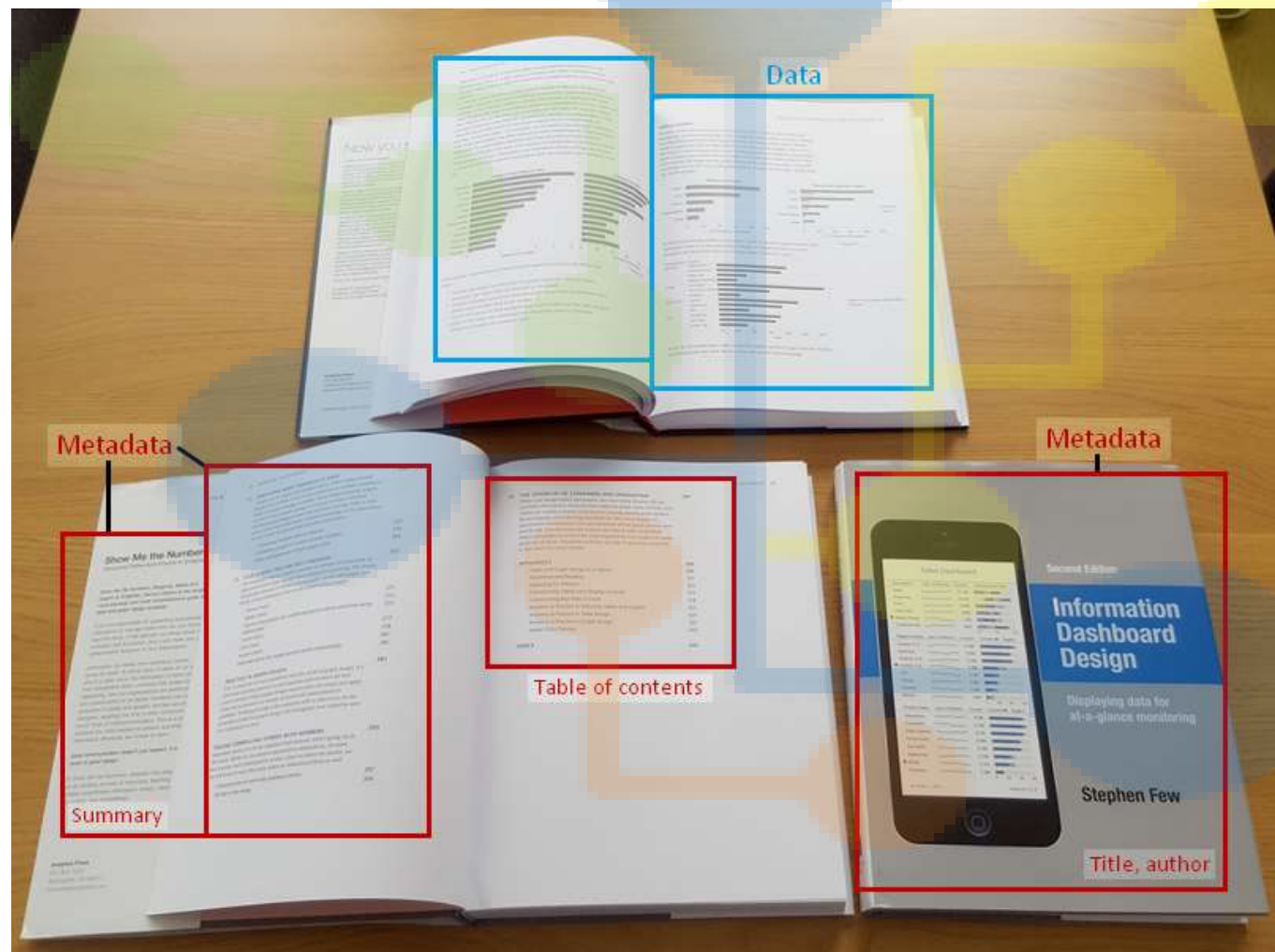


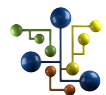
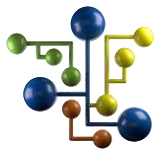
Em um livro, o que são dados e o que são metadados?





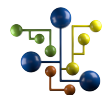
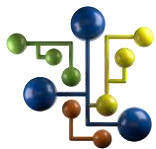
# O Que São Metadados?





# O Que é Master Data Management?



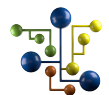
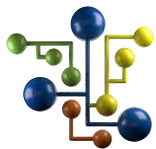


# O Que é Master Data Management?

## O Que é MDM?

O Master Data Management (MDM) é a tecnologia, as ferramentas e os processos que garantem que os dados principais sejam coordenados em toda a empresa. O MDM fornece um serviço de dados mestre unificado que fornece dados precisos, consistentes e completos em toda a empresa e para parceiros de negócios.



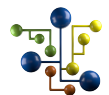
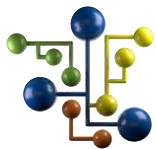


# O Que é Master Data Management?

## O Que é MDM?

O Gerenciamento de Dados Mestres (MDM) é o esforço feito por uma organização para criar uma única fonte de referência mestre para todos os dados críticos, levando a menos erros e menos redundância nos processos de negócio.





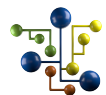
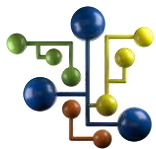
# O Que é Master Data Management?

Há algumas coisas que vale a pena observar nesta definição:

O MDM não é apenas um problema tecnológico. Em muitos casos, serão necessárias mudanças fundamentais no processo de negócio para manter dados mestres limpos e alguns dos problemas mais difíceis do MDM são mais políticos do que técnicos.





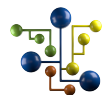
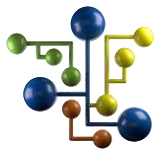


# O Que é Master Data Management?

Há algumas coisas que vale a pena observar nesta definição:

O MDM inclui a criação e a manutenção de dados mestres. Investir muito tempo, dinheiro e esforço na criação de um conjunto limpo e consistente de dados mestres é um esforço desperdiçado, a menos que a solução inclua ferramentas e processos para manter os dados mestres limpos e consistentes à medida que são atualizados e expandidos com o tempo.

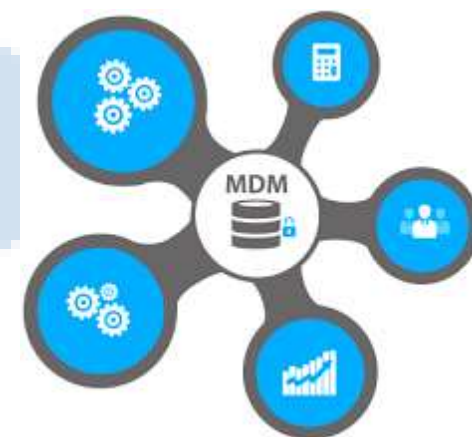


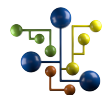
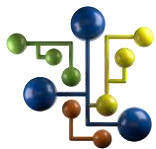


# O Que é Master Data Management?

## Por que MDM é Importante?

As soluções MDM abrangem uma ampla gama de práticas de limpeza, transformação e integração de dados. Conforme as fontes de dados são adicionadas ao sistema, o MDM inicia processos para identificar, coletar, transformar e reparar dados. Quando os dados atendem aos limites de qualidade, são criados esquemas e taxonomias para ajudar a manter uma referência mestre de alta qualidade. As organizações que usam o MDM têm a tranquilidade de que os dados em toda a empresa são precisos, atualizados e consistentes.



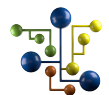
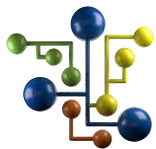


# O Que é Master Data Management?

## Quais os Benefícios do MDM?

Ao fornecer um ponto de referência para informações críticas, o MDM elimina redundâncias dispendiosas que ocorrem quando as organizações dependem de fontes de informações múltiplas e conflitantes. Por exemplo, o MDM pode certificar-se de que, quando as informações de contato do cliente forem alteradas, a organização não tentará vendas ou divulgação de marketing usando as informações antigas e novas. Ter múltiplas fontes de informação é um problema generalizado, especialmente em grandes organizações, e os custos associados podem ser muito altos.





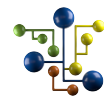
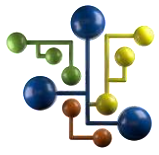
# O Que é Master Data Management?

## Quem Precisa de MDM?

O MDM é de particular interesse para grandes organizações, organizações com dados altamente distribuídos e aquelas que têm atividades de fusão e aquisição frequentes ou em larga escala. A aquisição de outra empresa cria desafios de integração de dados de grande alcance que o MDM foi projetado para mitigar. Assim, o MDM pode acelerar o tempo de retorno de uma aquisição.

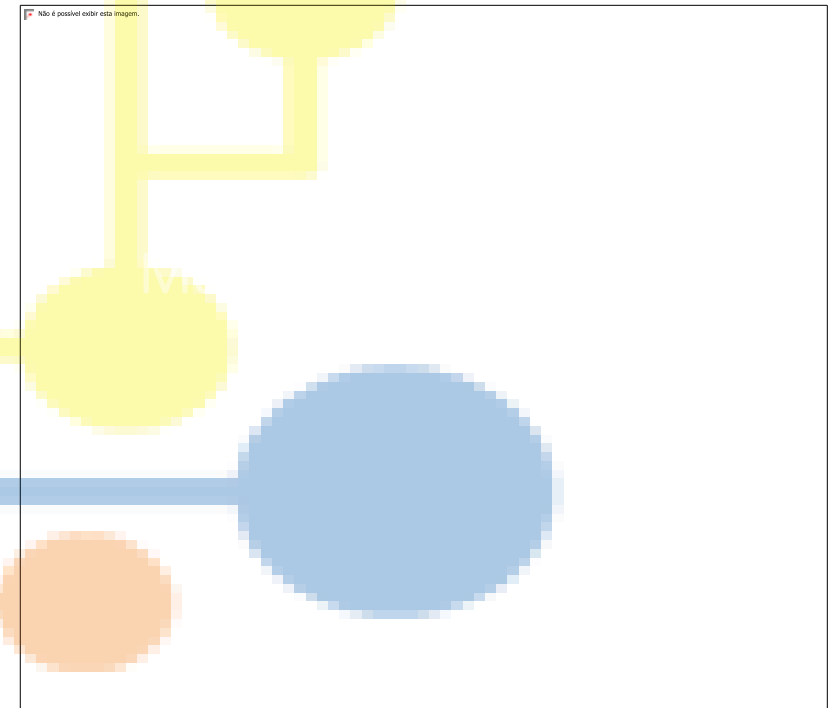
O MDM também ajuda empresas com linhas de produtos segmentadas, evitando experiências desintegradas dos clientes.



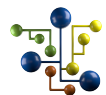
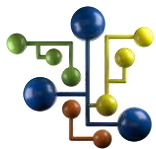


# Como Funciona o MDM na Prática

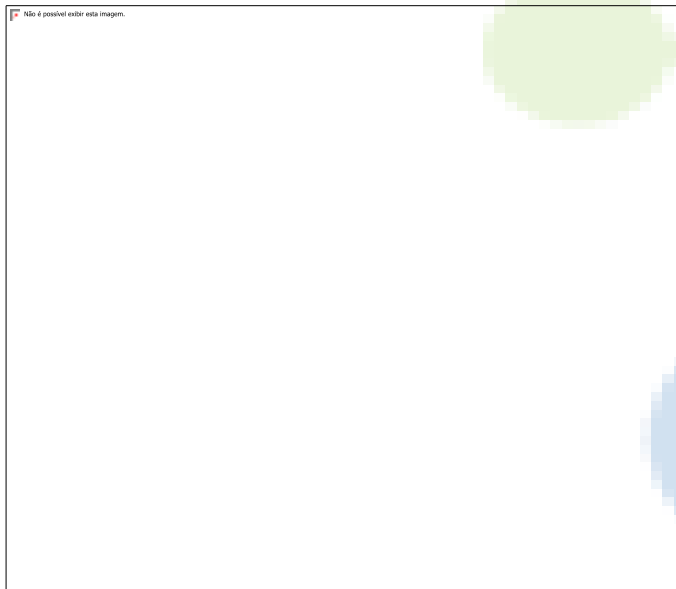
Temos 2 atividades principais em MDM:





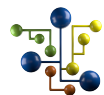
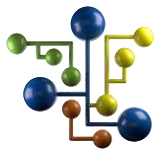


# Como Funciona o MDM na Prática



Quer você compre uma ferramenta MDM ou decida criar sua própria, existem duas etapas básicas para criar dados mestre:

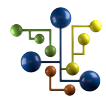
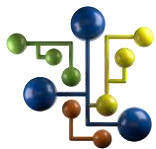
- Limpar e padronizar os dados
- Correspondência de dados de todas as fontes para consolidar duplicatas.



# Como Funciona o MDM na Prática

## Como Criamos a Lista Mestre?

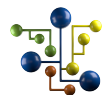
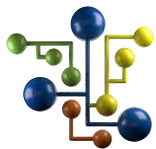
Antes de começar a limpar e normalizar os dados, você deve entender o modelo de dados para os dados mestre. Como parte do processo de modelagem, você deve ter definido o conteúdo de cada atributo e definido um mapeamento de cada sistema de origem para o modelo de dados mestre. Agora, você pode usar essas informações para definir as transformações necessárias para limpar os dados de origem.



# Como Funciona o MDM na Prática

## Como Criamos a Lista Mestre?

Limpar os dados e transformá-los no modelo de dados mestre é muito semelhante aos processos de Extração, Transformação e Carregamento (ETL) usados para preencher um Data Warehouse ou Data Lake. Se você já tiver ferramentas ETL e transformação definidas, talvez seja mais fácil modificá-las conforme necessário para os dados mestres, em vez de aprender uma nova ferramenta.

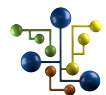
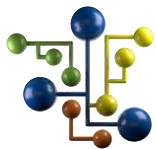


# Como Funciona o MDM na Prática

## Como Criamos a Lista Mestre?

Aqui estão algumas funções típicas de limpeza de dados:

- Normalizar os formatos de dados
- Substituir os valores ausentes
- Padronizar valores
- Eliminar valores duplicados
- Mapeamento de Atributos

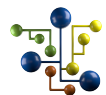
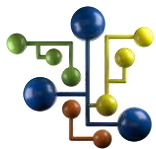


# Como Funciona o MDM na Prática

E Como Podemos Manter a Lista Mestre?



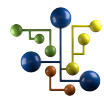
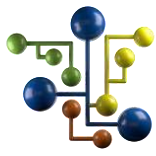




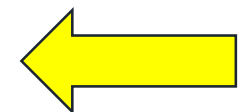
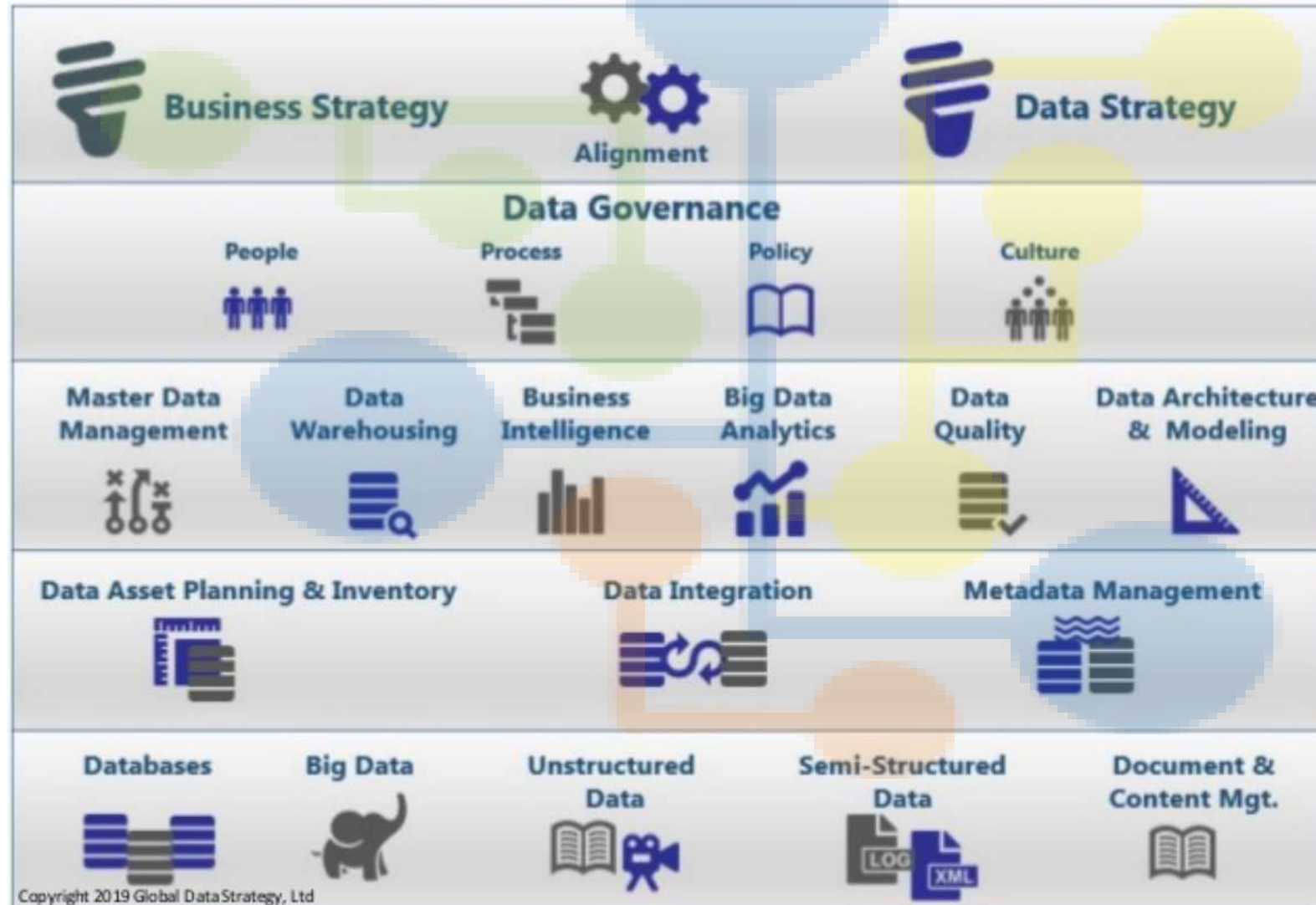
# Como Funciona o MDM na Prática

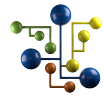
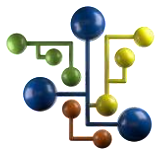
Existem várias formas de manter a lista mestre:

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



# Diferença Entre Gerenciamento de Documentos e Gerenciamento de Conteúdo



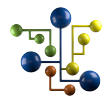
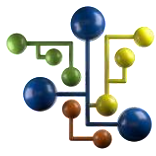


# Diferença Entre Gerenciamento de Documentos e Gerenciamento de Conteúdo

Muitas empresas buscam otimizar os processos de negócios automatizando o gerenciamento de informações comerciais.

No entanto, o número de soluções de gerenciamento pode ser confuso, especialmente quando os limites entre as soluções não estão bem definidos.

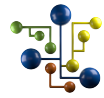
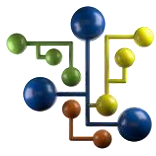
Por exemplo, considere Sistemas de Gerenciamento de Documentos (DMS) e Sistemas de Gerenciamento de Conteúdo (CMS): algumas pessoas usam os termos DMS e CMS como sinônimos, mas isso não é totalmente correto - embora eles tenham alguns recursos semelhantes.



# Diferença Entre Gerenciamento de Documentos e Gerenciamento de Conteúdo

Em geral, um 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 eletrônicas. Ele também suporta controle de versão, colaboração e fluxos de trabalho.

Sistemas CMS são usados para criar e gerenciar vários tipos de conteúdo digital, além de documentos no sentido tradicional. Por exemplo, além de ser usado para PDFs, arquivos do Word ou arquivos do Excel, um CMS pode ser usado para gerenciar coisas como imagens, páginas web, registros e arquivos flash.

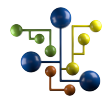
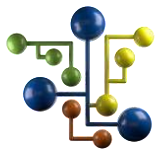


# Diferença Entre Gerenciamento de Documentos e Gerenciamento de Conteúdo

Como o CMS e o DMS diferem!

Apesar das semelhanças, as ferramentas DMS e CMS diferem de várias maneiras. Cada um tem alguns recursos importantes que o outro não tem e lidam com diferentes tipos de dados. Aqui estão alguns exemplos:

Um DMS gerencia dados estruturados e é focado em documentos no sentido tradicional em formatos como Word, PDF, PowerPoint, Excel, etc. Um CMS, por outro lado, pode manipular dados estruturados e dados não estruturados, como conteúdo da Web (Arquivos HTML e PDF) e ativos digitais (imagens e arquivos de áudio e vídeo).



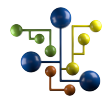
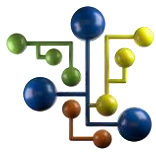
# Diferença Entre Gerenciamento de Documentos e Gerenciamento de Conteúdo

Como o CMS e o DMS diferem!

Apesar das semelhanças, as ferramentas DMS e CMS diferem de várias maneiras. Cada um tem alguns recursos importantes que o outro não tem e lidam com diferentes tipos de dados. Aqui estão alguns exemplos:

**Os principais propósitos de um DMS são conformidade normativa e gerenciamento de fluxo de trabalho, enquanto os principais propósitos de um CMS são armazenamento, recuperação e publicação de conteúdo.**



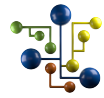
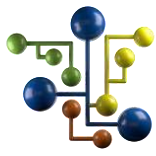


# Diferença Entre Gerenciamento de Documentos e Gerenciamento de Conteúdo

Como o CMS e o DMS diferem!

Apesar das semelhanças, as ferramentas DMS e CMS diferem de várias maneiras. Cada um tem alguns recursos importantes que o outro não tem e lidam com diferentes tipos de dados. Aqui estão alguns exemplos:

**Os aplicativos DMS possuem recursos avançados de imagem e digitalização, como reconhecimento óptico de caracteres (OCR), reconhecimento de caracteres de impressão digital (HCR), reconhecimento de marcas ópticas (OMR) e muito mais. As ferramentas do CMS geralmente não suportam essas funções.**

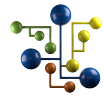
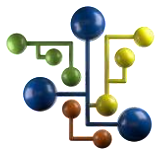


# Diferença Entre Gerenciamento de Documentos e Gerenciamento de Conteúdo

Como o CMS e o DMS diferem!

Apesar das semelhanças, as ferramentas DMS e CMS diferem de várias maneiras. Cada um tem alguns recursos importantes que o outro não tem e lidam com diferentes tipos de dados. Aqui estão alguns exemplos:

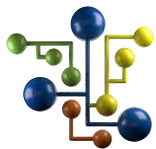
**A integração com sistemas corporativos (como planejamento de recursos corporativos e ferramentas de gerenciamento de relacionamento com clientes) é essencial para um DMS, mas secundário para um CMS.**



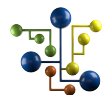
# Governança de Dados e Big Data

Big Data pode ser definido pelos seus 4 V's: Volume, Variedade, Velocidade e Veracidade.

Com base nessa definição, qual seria um dos aspectos técnicos mais relevantes em Governança de Big Data?



Data Science  
Academy



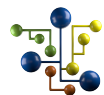
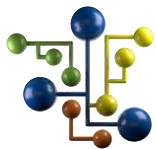
Data Science  
Academy

Data Science Academy jaquelinefernandes6@gmail.com 5e84edfee32fc3397c122ab7

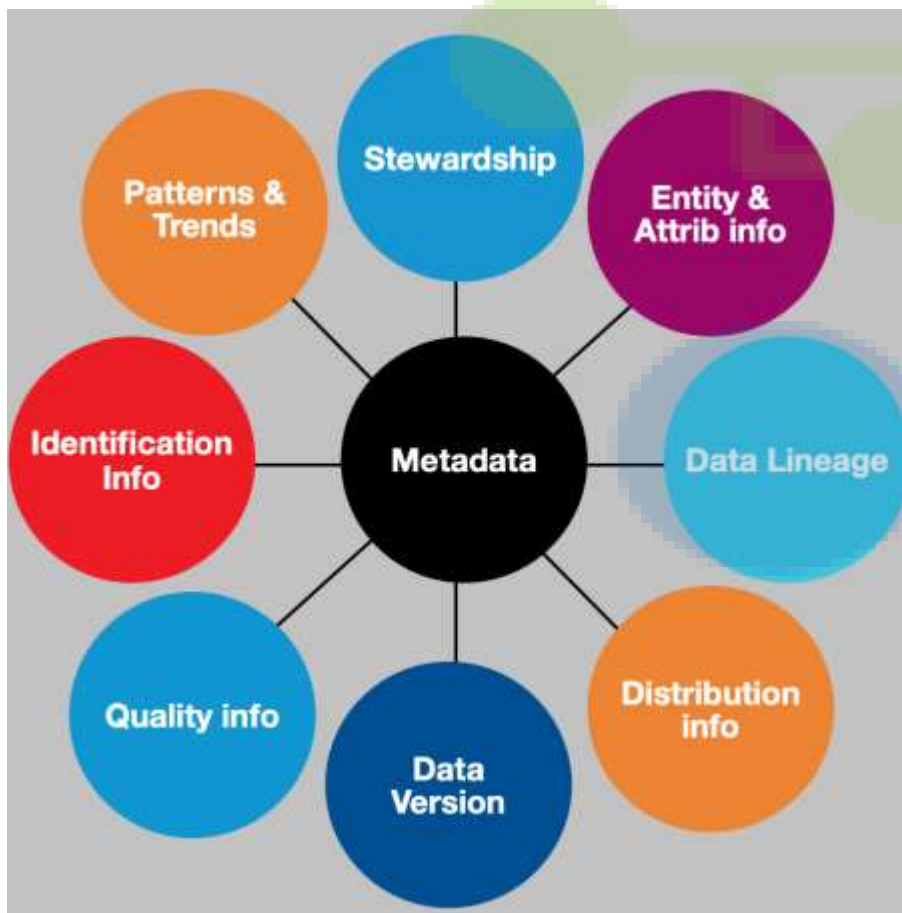
# Governança de Dados e Big Data

An abstract network diagram with several large circular nodes connected by lines. The nodes are colored in shades of blue, green, yellow, and orange. The word 'METADADOS' is overlaid in the center of the diagram.

**METADADOS**

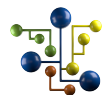
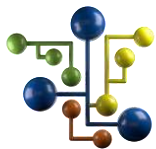


# Governança de Dados e Big Data

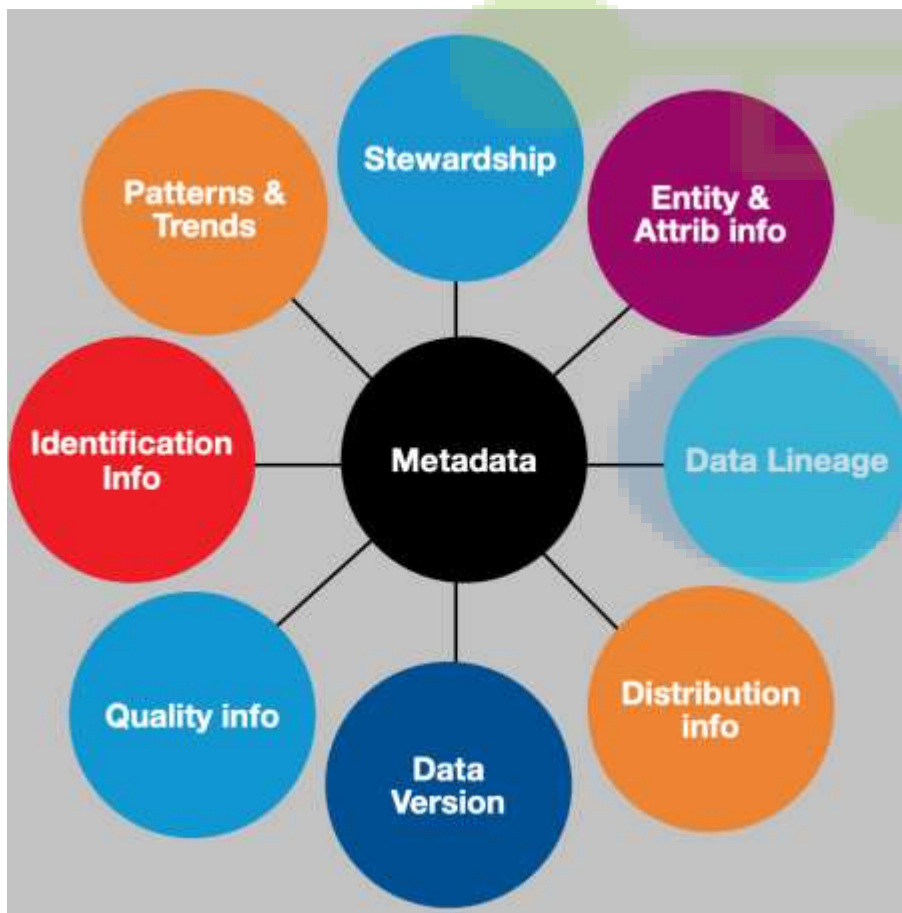


O Data Lake armazena grande quantidade de dados estruturados e não estruturados em diferentes camadas.

Enquanto os dados estão crescendo para terabytes e petabytes e o Data Lake está sendo usado pela empresa, é provável que você se depare com perguntas/desafios, como 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.

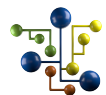
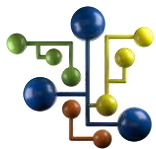


# Governança de Dados e Big Data



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



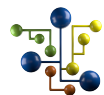
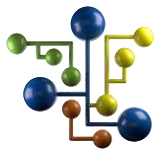


# Governança de Dados e Big Data

Benefícios e Funções de Camada de Metadados na Governança de Big Data:

- A camada de metadados fornece informações significativas sobre a importância dos dados armazenados no Data Lake para seus usuários. Isso é realizado identificando de forma inteligente cada bit de dados à medida que é ingerido. Uma camada de metadados bem construída permitirá a organização aproveitar 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)
- Aprendizado de Máquina como Serviço
- Provisionamento de dados (DP)



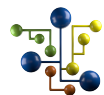
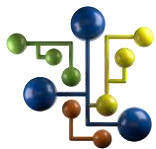


# Governança de Dados e Big Data

## Benefícios e Funções de Camada de Metadados na Governança de Big Data:

- A camada de metadados define a estrutura dos arquivos no “Raw Zone” e descreve as entidades no lado do arquivo. Usando a descrição do nível básico, a evolução do esquema do arquivo ou registro é rastreado por um esquema de versão. Isso vai eventualmente, permitir que você crie associação entre várias entidades e, assim, facilitar a navegação e pesquisa dos dados que o usuário final está procurando.



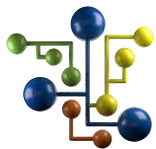


# Governança de Dados e Big Data

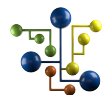
Benefícios e Funções de Camada de Metadados na Governança de Big Data:

- Na camada de consumo, é muito conveniente saber a fonte dos dados ao passar por um relatório pois pode haver uma versão diferente dos dados de entrada. A clareza dos relacionamentos nos metadados ajuda a resolver ambiguidades e inconsistências na determinação de associações entre entidades armazenadas em todos os dados.





Data Science  
Academy

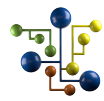
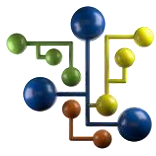


Data Science  
Academy

Data Science Academy jaquelinefernandes6@gmail.com 5e84edfee32fc3397c122ab7

# Governança de Dados com Apache Atlas

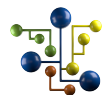
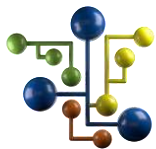




# Governança de Dados com Apache Atlas

Apache Atlas é uma iniciativa de Governança de Dados da Hortonworks em Cluster Hadoop. Foi inicialmente criado a partir do Hortonworks e depois levado para o Apache como um projeto de nível superior.

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.

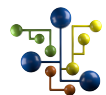
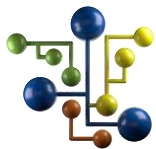


# Governança de Dados com Apache Atlas

Principais Características e Funcionalidades do Apache Atlas:

## Classificação de 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 processos de origem, destino e derivação.
- Exportar metadados para sistemas de terceiros.



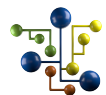
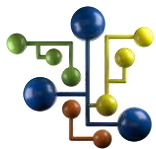
# Governança de Dados com Apache Atlas

Principais Características e Funcionalidades do Apache Atlas:

## 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.



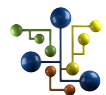
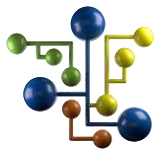


# Governança de Dados com Apache Atlas

Principais Características e Funcionalidades do Apache Atlas:

## Pesquisa e Linhagem

- Caminhos de navegação predefinidos para explorar os dados de classificação e auditoria.
- Os recursos de pesquisa baseados em texto localizam dados relevantes e evento de auditoria no Data Lake com rapidez e precisão.
- Permite navegar pela visualização da linhagem do conjunto de dados, permitindo que os usuários possam aplicar drill-down em informações operacionais, de segurança e procedência relacionadas.

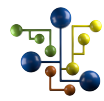
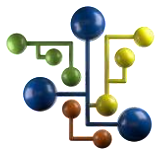


# Governança de Dados com Apache Atlas

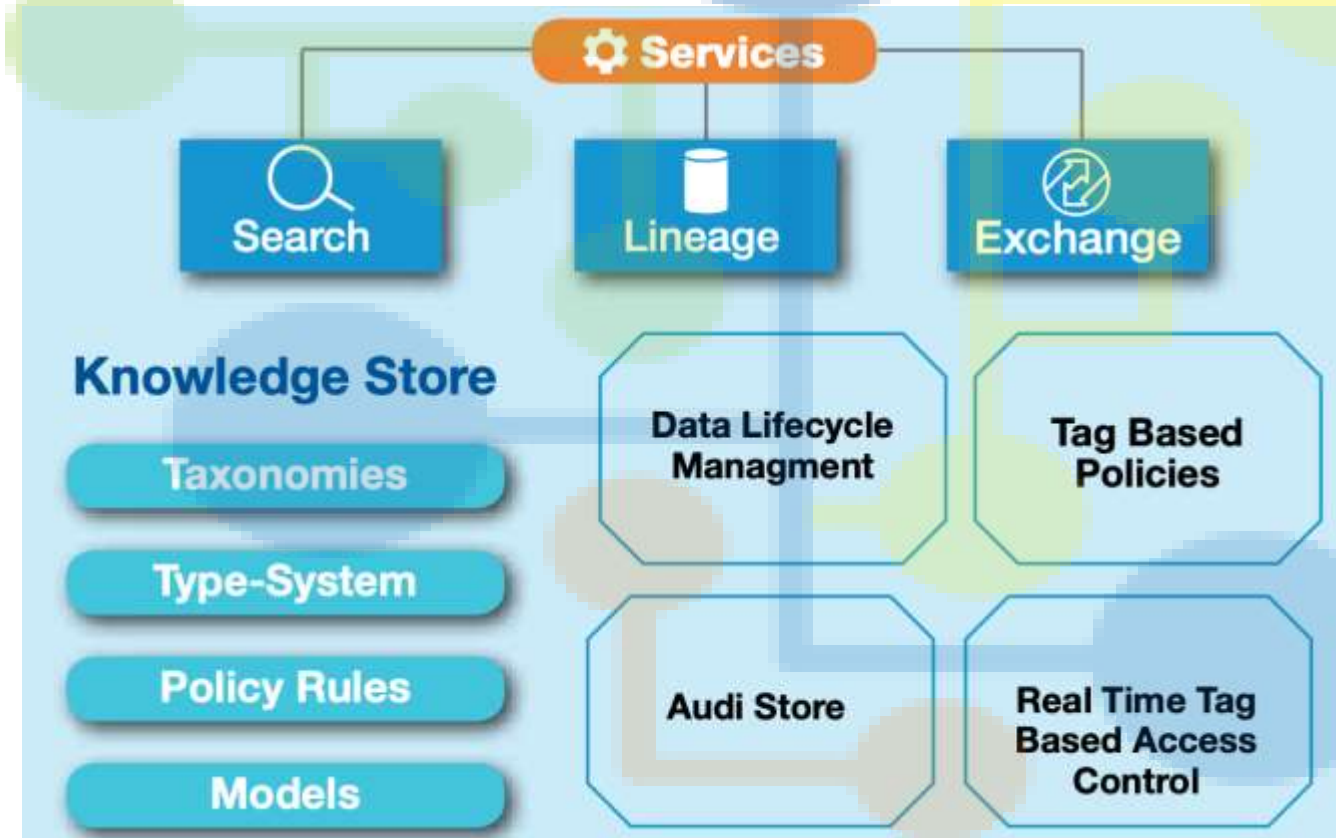
Principais Características e Funcionalidades do Apache Atlas:

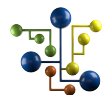
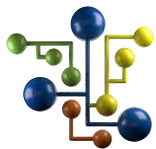
## Mecanismo de Segurança e Políticas

- 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 (ou seja, reidentificação).
- Mascaramento de nível de coluna e linha com base nos valores da célula e atributos.



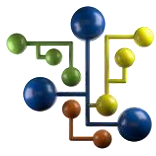
# Governança de Dados com Apache Atlas



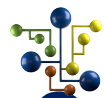


# Governança de Dados com Apache Atlas

O Apache Atlas pode ser encontrado como um módulo do curso Segurança e Alta Alta Disponibilidade de Dados, da Formação Engenheiro de Dados, aqui na DSA.



Data Science  
Academy



Data Science  
Academy

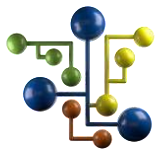
Data Science Academy jaquelinefernandes6@gmail.com 5e84edfee32fc3397c122ab7

# Ó Que Você Deve Saber Sobre a Lei de Proteção aos Dados

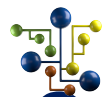
**The EU General Data Protection Regulation (GDPR) is the most important change in data privacy regulation in 20 years.**

The regulation will fundamentally reshape the way in which data is handled across every sector, from healthcare to banking and beyond.

<https://eugdpr.org/>



Data Science  
Academy



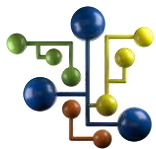
Data Science  
Academy

Data Science Academy jaquelinefernandes6@gmail.com 5e84edfee32fc3397c122ab7

# Ó Que Você Deve Saber Sobre a Lei de Proteção aos Dados







Data Science  
Academy



Data Science  
Academy

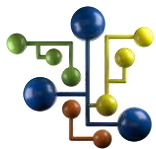
Data Science Academy jaquelinefernandes6@gmail.com 5e84edfee32fc3397c122ab7

# O Que Você Deve Saber Sobre a Lei de Proteção aos Dados

O que é a Lei Geral de Proteção de Dados Pessoais, então?







Data Science  
Academy



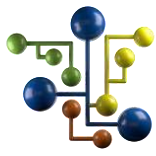
Data Science  
Academy

Data Science Academy jaquelinefernandes6@gmail.com 5e84edfee32fc3397c122ab7

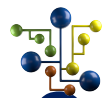
# Ó Que Você Deve Saber Sobre a Lei de Proteção aos Dados

Como os dados pessoais deverão ser coletados e tratados?





Data Science  
Academy



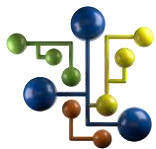
Data Science  
Academy

Data Science Academy jaquelinefernandes6@gmail.com 5e84edfee32fc3397c122ab7

# O Que Você Deve Saber Sobre a Lei de Proteção aos Dados

O que acontece em caso de vazamento de dados?





Data Science  
Academy



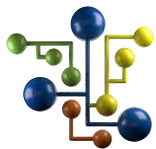
Data Science  
Academy

Data Science Academy jaquelinefernandes6@gmail.com 5e84edfee32fc3397c122ab7

# Ó Que Você Deve Saber Sobre a Lei de Proteção aos Dados

Qual a punição para descumprimento da lei?





Data Science  
Academy



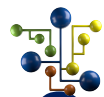
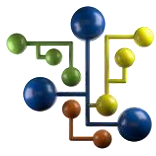
Data Science  
Academy

Data Science Academy jaquelinefernandes6@gmail.com 5e84edfee32fc3397c122ab7

# Ó Que Você Deve Saber Sobre a Lei de Proteção aos Dados

Vale só para empresas brasileiras ou estrangeiras também?



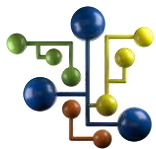


# O Que Você Deve Saber Sobre a Lei de Proteção aos Dados

Quem vai fiscalizar?







Data Science  
Academy



Data Science  
Academy

Data Science Academy jaquelinefernandes6@gmail.com 5e84edfee32fc3397c122ab7

# Ó Que Você Deve Saber Sobre a Lei de Proteção aos Dados

Quando a Lei Geral de Proteção de Dados Pessoais começa a valer?

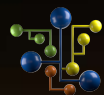
Começo de 2020



# obrigado

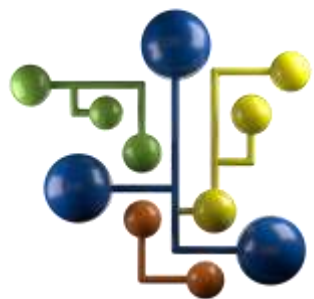






# Muito Obrigado!

## Continue trilhando uma excelente jornada de aprendizagem!



Governança de Dados

