

GOVERNANÇA DE DADOS

Implementando uma estratégia de sucesso

O guia definitivo de como estruturar um framework focado em segurança e resultados



Antes de começar sua leitura

Esse e-book é **interativo**. Sempre que houver um texto ou área clicável você será redirecionado para uma área específica.

Esperamos que goste do conteúdo e que isso te ajude a ter uma leitura melhor!

Índice

O que é Governança de Dados e por que sua empresa precisa disso? 04

Investir em governança de dados é necessário se você busca: 05

- Uma visão consistente e segura das informações que a empresa capta 05
- Fortalecimento do compliance 05
- Qualidade de dados 06
- Equilíbrio perfeito entre automação e gerenciamento manual 06

O Ciclo perfeito da Governança de Dados 07

A base de tudo 08

- PESSOAS 09
- PROCESSOS 10
- TECNOLOGIA 11

Entendendo as etapas da Governança de Dados 12

- 1 | Propriedade 13
- 2 | Acessibilidade 15
- 3 | Segurança 17
- 4 | Qualidade 19
- 5 | Aprendizado 20

Encontrando o framework perfeito para a sua implementação 21

- O antigo e insustentável modelo de governança de dados 22
- O framework mais comum e sua luta para centralizar os dados em um só local 24

Framework colaborativo: o futuro da governança de dados 26

Mão na massa: o passo a passo da implementação 28

- Entendendo o cenário atual 29
- Aplicando o Data Profiling 30
- Escalando o acesso a dados 31
- Protegendo e mascarando os dados 32
- Facilitando o acesso para todas as áreas da empresa 33

Potencializando a operação com o Talend 34

- Criando uma cultura que valoriza a governança de dados 35
- Construindo um data lake completo para permitir uma boa estratégia de governança 35

Conclusão 37

O que é Governança de Dados e por que sua empresa precisa disso?



Dados realmente potencializam tudo o que fazemos

Jeff Weiner, CEO do LinkedIn

Você já deve ter lido em muitos lugares e, todos nós já sabemos, que os dados são o novo petróleo. O que muitos esquecem é de focar em como trabalhar esses dados de maneira eficiente. Por esse motivo, criamos esse material: mostrar como uma boa estratégia de governança de dados pode realmente mudar o rumo de uma empresa.

A Governança de Dados é um conjunto de processos, políticas, padrões e métricas que garantem um uso eficiente das informações para atingir seus objetivos de negócio.

Sem esse conjunto, sua empresa estará perdendo tempo e dinheiro. Não podemos só trabalhar os dados, mas sim fazer com que eles trabalhem para nós. Há alguns pontos cruciais para manter o uso de dados eficiente para trazer mais e mais resultados. Veja a seguir.

Investir em governança de dados é necessário se você busca:

- Uma visão consistente e segura das informações que a empresa capta

O data governance proporciona um alinhamento maior dos dados da empresa, gerando padrões facilmente identificáveis e de acesso simplificado para todos os colaboradores da empresa.

Não significa que todos terão acesso a todos os dados, mas sim as informações necessárias para cada área. Essa distribuição mais democrática alavanca as tomadas de decisão de cada setor, tanto individualmente quanto no coletivo da companhia.

- Fortalecimento do compliance

É possível ampliar o acesso a dados de forma segura? A resposta é sim. Quando as informações são cruzadas da maneira correta, as pessoas terão acesso a informações que realmente fazem sentido para o trabalho delas.

Uma estratégia sólida deve estar perfeitamente alinhada com as leis de proteção de dados, como a LGPD. Isso não significa restringir o acesso, mas sim distribuí-lo de forma inteligente.

Investir em governança de dados é necessário se você busca:

- Qualidade de dados

O data quality também está diretamente relacionado com a governança de dados, uma vez que os padrões que citamos anteriormente contribuem para a qualificação das informações.

Isso acontece pois, ao termos os dados padronizados, identificar falhas é uma tarefa muito mais simples. Uma boa governança proporciona dados completos, consistentes e precisos.

- Equilíbrio perfeito entre automação e gerenciamento manual

O mundo dos dados é altamente automatizado e robotizado. Uma boa estratégia de governança traz o melhor dos mundos, estabelecendo códigos de conduta e boas práticas para a manipulação de dados.

Todo o sistema trabalha com a velocidade de uma máquina e a precisão humana. O foco é trazer resultados que realmente façam sentido para toda a operação.

O Ciclo perfeito da Governança de Dados





O Ciclo perfeito da Governança de Dados

A base de tudo

Antes de entender o ciclo da Governança de Dados, precisamos analisar a sua base, que é composta por pessoas, processos e tecnologia. Sem esses três pontos será impossível definir um framework saudável para a empresa.



O Ciclo perfeito da Governança de Dados

PESSOAS

Dados precisam ser democráticos

Ter profissionais altamente qualificados nesse processo é essencial, mas não é apenas aí que o pilar das pessoas é baseado. O conceito aborda a democratização dos dados para os diferentes setores da companhia.

Significa que, tanto o setor financeiro quanto o setor de marketing, por exemplo, devem ter acesso a dados que sejam valiosos para entender os cenários e ajudar no desenvolvimento das áreas. Outro ponto relacionado é a colaboração de pessoas de fora da área de tecnologia na qualificação e manipulação dos dados.



PROCESSOS

Como fazer é tão importante quanto o que fazer

Ter processos bem definidos é o sonho de todo gestor. No contexto de dados esse sonho é mais importante ainda. As boas práticas e guidelines são a verdadeira bússola para que as informações delicadas não corram riscos.

Além do fator humano, é possível usar machine learning e outros tipos de automação para restringir o acesso à dados para cada colaborador da empresa. Como já dissemos, cada setor depende de um tipo de informação, por isso filtrar é extremamente importante para a segurança e performance.



O Ciclo perfeito da Governança de Dados

TECNOLOGIA

A melhor forma de acelerar resultados

É impossível trabalhar dados sem o auxílio da tecnologia. Utilizar ela da melhor forma é o que diferencia as grandes empresas das gigantes do mercado. Escolher ferramentas, automatizar tarefas e acompanhar resultados é essencial para o crescimento.

Com a ferramenta correta você consegue centralizar grande parte da operação e ganhar muito mais agilidade nas ações e controle do seu framework de Data Governance.

Entendendo as etapas da Governança de Dados

Por mais que seja uma estratégia cílica, é preciso ter muita atenção em cada etapa para que tudo corra como o esperado. Cada ponto tem sua especificidade e precisa ser trabalhado da melhor forma.

Nunca se esqueça: data governance é sobre o caminho, não o destino. Todos os parâmetros devem ser constantemente revisados e testados para melhorar ainda mais a performance da estratégia.

Mas antes de colocar as mãos na massa, vamos entender cada etapa desse processo para colocá-lo em prática da melhor forma possível.





O Ciclo perfeito da Governança de Dados

1 | PROPRIEDADE

Antes de falar sobre a importância da propriedade dos dados é preciso entender dois conceitos básicos: Data Owners e Data Stewards.

Essas duas peças são essenciais para uma boa governança de dados, uma vez que elas definem quais são os níveis de acesso e funções dentro de cada área.

O Data Owner normalmente é mais sênior, enquanto os Data Stewards são os profissionais de cargo mais júnior. O primeiro, é responsável pela questão mais gerencial e o segundo pelos detalhes técnicos de qualificação e registro dos dados.



O Ciclo perfeito da Governança de Dados

Por exemplo, no setor financeiro é comum que o CFO seja o Data Owner, enquanto outros funcionários são os Data Stewards. Nesse cenário, o Owner é responsável por resolver problemas de qualidade de dados, enquanto os Stewards analisam os problemas e indicam correções e possíveis mudanças de rota.

É claro que essa questão se adapta em cada tipo de empresa, uma vez que em empresas menores é possível atribuir as funções de owner e steward para apenas uma pessoa. O que vai ditar essa prática é o porte da companhia e o volume de dados que é gerado.

O ciclo começa em propriedade de dados por esse motivo. É preciso entender quem são os proprietários desses dados e como eles serão trabalhados da melhor forma, entregando as informações de forma segura e democrática.



O Ciclo perfeito da Governança de Dados

2 | ACESSIBILIDADE

Das ferramentas de BI ao CRM, tudo precisa ser alimentado com dados para performar bem. Se a companhia busca resultados incríveis, é preciso de um norte para cada tipo de ação que visa o crescimento.

Os dados são a bússola para essa busca. Promover a real transformação digital sem a ajuda de dados é quase impossível. Por isso a acessibilidade é vital para a governança de dados.

Para isso é preciso ter certeza de que a empresa está ativamente coletando dados de qualidade. Todo o processo de coleta, limpeza e qualificação das informações deve ser milimetricamente calculado, tornando os dados uma fonte segura para a tomada de decisão.



O Ciclo perfeito da Governança de Dados

Em um segundo momento é preciso garantir que, quem precisa de informações, está tendo acesso a elas. Para isso é utilizado o Talend Data Catalog, uma frente da ferramenta focada em governança de dados.

Em uma empresa do ramo da saúde, o setor financeiro não precisa saber da enfermidade de um paciente, apenas as informações cadastrais referentes à pagamento e cobrança, por exemplo.

Isso facilita o dia a dia de cada colaborador, indo direto ao ponto que é necessário para sua atividade dentro da companhia. Outro fator importantíssimo para a estratégia é a facilidade de acesso. Com dashboards e interfaces mais amigáveis, o colaborador que não é do TI consegue entender de forma muito mais clara cada tipo de informação, aumentando drasticamente a produtividade.



O Ciclo perfeito da Governança de Dados

3 | SEGURANÇA

Podemos dizer que todas as empresas lidam com dados e sabemos o quanto importante é manter essas informações seguras. Quando lidamos com big data, a responsabilidade aumenta infinitamente.

Por isso que a segurança é um dos aspectos mais importantes da governança de dados. Em um panorama mundial de crescimento acelerado da tecnologia, manter os dados seguros é essencial para garantir o crescimento e não se envolver em polêmicas de vazamentos de dados ou ataques cibernéticos.

Para combater isso é preciso usar a própria tecnologia. Estabelecer níveis de acesso e aproveitar ao máximo o machine learning é o melhor caminho para ter dados confiáveis, organizados, acessíveis e principalmente seguros.



O Ciclo perfeito da Governança de Dados

Até mesmo a esfera jurídica está se movimentando para aumentar o compliance e assegurar o direito do cidadão. A LGPD (Lei Geral de Proteção de Dados) é um conjunto de regras para armazenamento, coleta e processamento de dados pessoais feito por empresas privadas e até mesmo pelo poder público.

Quando se trata do uso de dados para fins comerciais, ela fiscaliza as empresas e como elas estão utilizando as informações. Pisar em falso com relação a isso pode gerar consequências gravíssimas, envolvendo multas de até 2% do faturamento da empresa.



O Ciclo perfeito da Governança de Dados

4 | QUALIDADE

Conduzir uma empresa com dados desqualificados é perigosíssimo. No início as estratégias podem parecer dar certo, mas a médio e longo prazo é um verdadeiro tiro no pé.

Não investir na qualificação dos dados gera falsas impressões de controle e conhecimento do mercado em que você está inserido. Como já dissemos em nosso e-book sobre o Talend: reparar erros causados por dados ruins pode ser até 100x mais caro quando comparado com a qualificação precoce no processo de ETL.

Ainda no espectro do Data Quality, agilidade também é essencial. Mas como qualificar dados da maneira correta e ainda por cima ser rápido? A resposta está na tecnologia que é adotada.

Ferramentas como o Talend tornam o processo de extração de dados até 30x mais rápido. Enquanto 10 informações são qualificadas manualmente, 300 são qualificadas no mesmo período de tempo se há uma boa ferramenta de ETL envolvida.



O Ciclo perfeito da Governança de Dados

5 | APRENDIZADO

Agora que todos os parâmetros de propriedade, acessibilidade, segurança e qualidade estão estabelecidos, é preciso voltar alguns passos e analisar tudo o que foi feito até o momento.

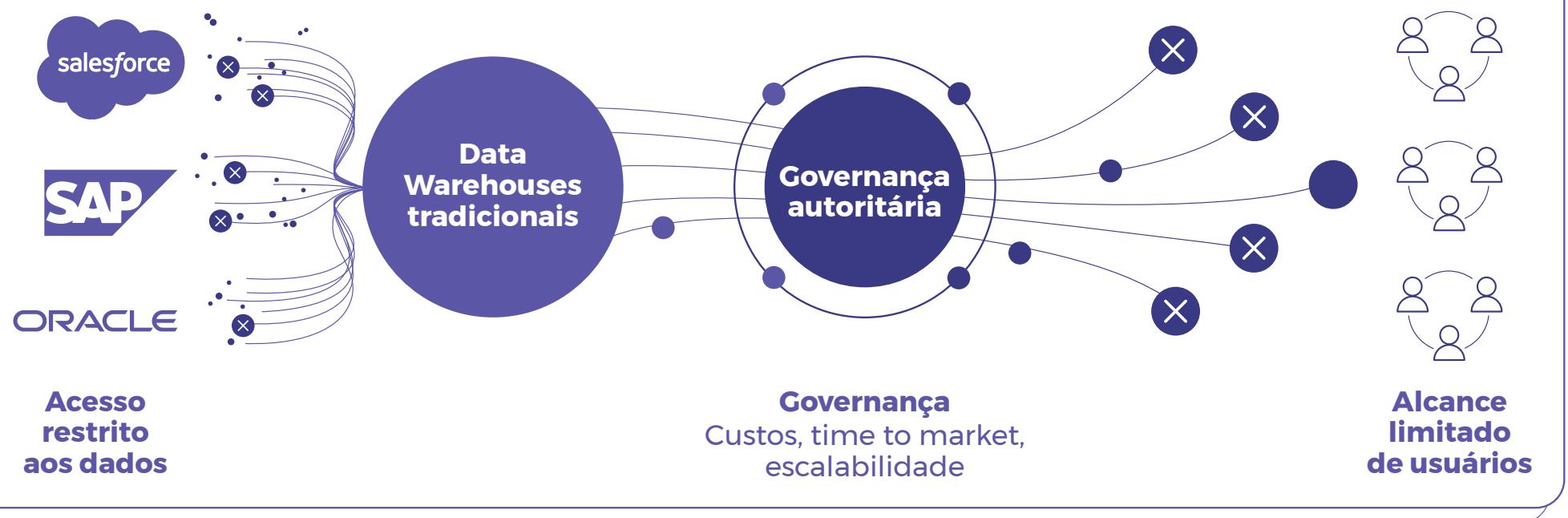
A etapa de aprendizado é essencial na análise da operação como um todo, identificando gargalos e possibilidades de crescimento para a empresa. É nesse momento que problemas são encontrados na fase inicial, economizando muito mais recursos da companhia.

Por mais que muitos problemas já tenham sido encontrados e resolvidos durante a implementação do projeto, é preciso ter um estudo recorrente de tudo o que está sendo feito e será feito nos próximos passos. Jamais negligencie a importância da etapa de aprendizado se você busca resultados sólidos e constantes.

Se possível, reserve um tempo dos heads da operação para olhar cada passo com calma e estudar o que será melhorado nas próximas etapas.

Encontrando o framework perfeito para a sua implementação

O antigo e insustentável modelo de governança de dados

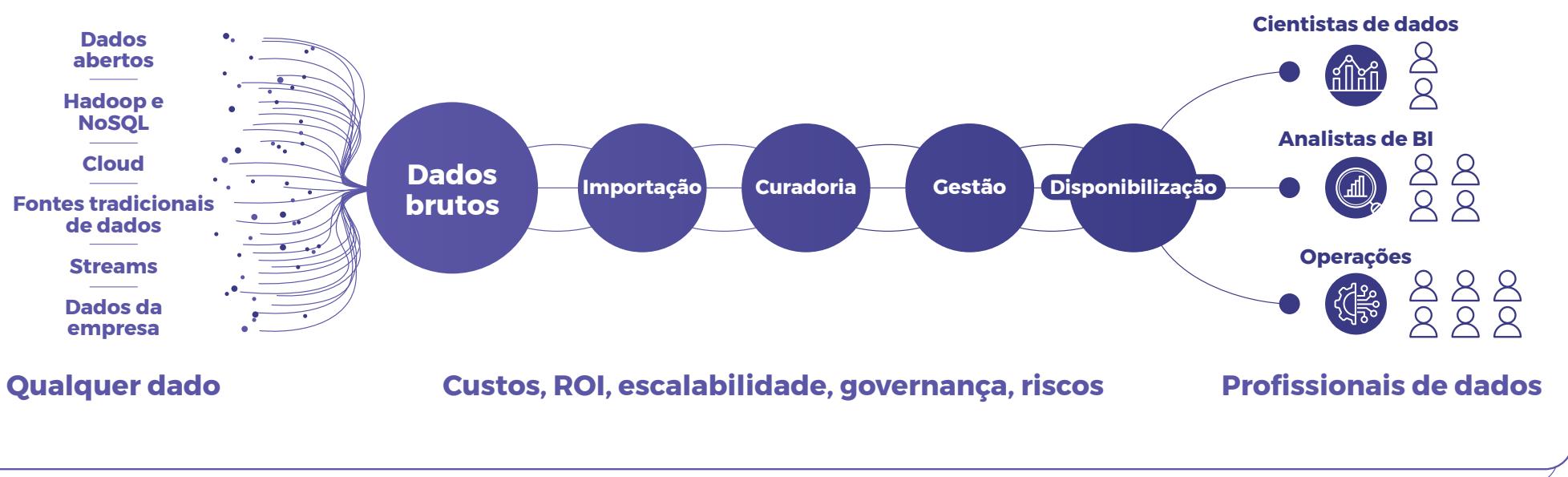


O modelo tradicional de data governance não faz mais sentido para empresas que buscam escala na nova era digital. Antigamente pouquíssimas pessoas tinham acesso aos dados, que também eram poucos, criando uma forte centralização das fontes de dados que, são fáceis de gerenciar, porém não trazem a amostra necessária para colher bons frutos e tomar decisões seguras.

Para esse dado ser consumido em análises era preciso passar por um data warehouse em um processo antiquado e lento, analisado para ser qualificado, remodelado para um data mart e enfim remodelado mais uma vez para ser usado em uma ferramenta de BI para chegar ao relatório.

Imagine fazer esse processo com milhões de dados em tempo recorde para não perder o timing do mercado. Seria impossível gerar escala em grandes volumes. Hoje em dia é preciso descentralizar o acesso a informações valiosas e democratizá-lo da melhor forma possível.

O framework mais comum e sua luta para centralizar os dados em um só local

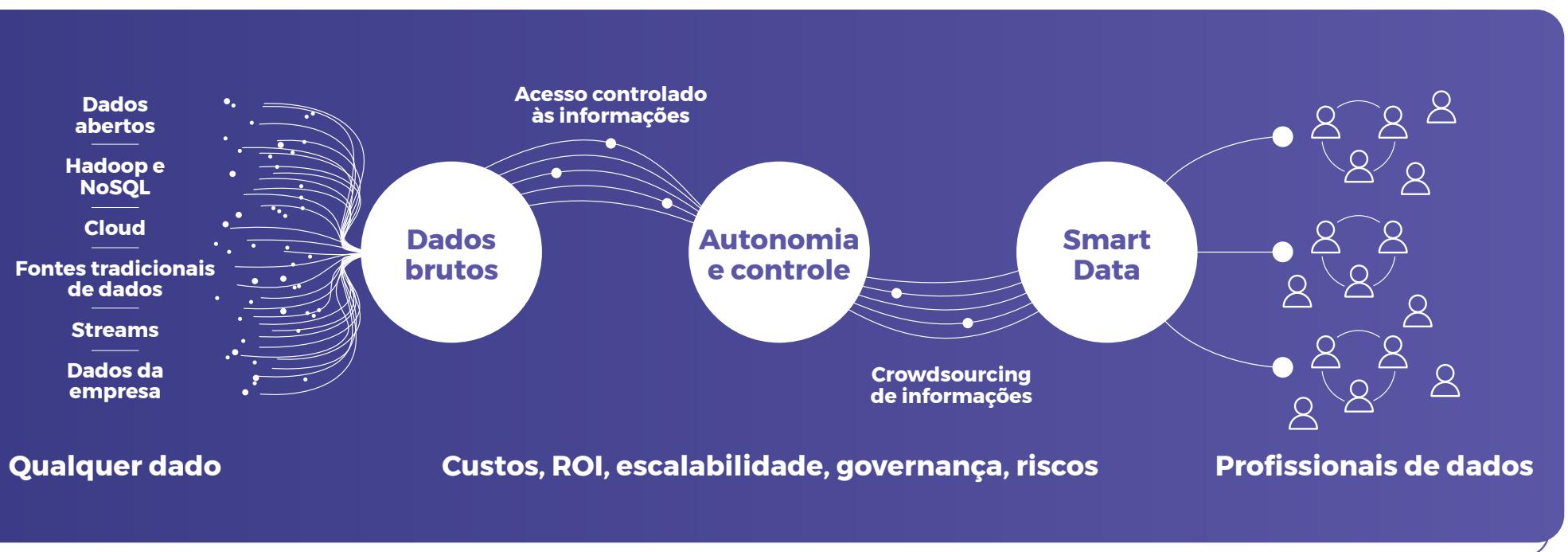


Com a chegada do Big Data o crescimento da cultura ágil foi inevitável. A necessidade de ter escala e um processamento mais rápido bateu à porta das grandes empresas, surgindo assim os data lakes. Enquanto o modelo antigo trazia uma visão de modelar os dados e então depois torná-los legíveis para os colaboradores com acesso limitadíssimo, essa nova prática foi totalmente no caminho inverso.

Os dados chegavam brutos ao data lake e então eram estruturados com base no data quality e as regras de segurança e governança de dados estabelecidas pela empresa antes de serem recebidos por um grupo (ainda seletivo, porém mais abrangente) de colaboradores.

Esse tipo de estratégia ainda funciona para muitos tipos de empresa, porém, quanto mais acelerado é o crescimento, maior é a dificuldade de centralização dos dados. O grande aumento das fontes de informação gerou e ainda gera dores de cabeça para os gestores que precisam controlar e organizar todos os dados em apenas um repositório.

Framework colaborativo: o futuro da governança de dados



O que não foi percebido no segundo modelo é que as fontes de dados aumentam diariamente e isso não só pode como em algum momento vai dificultar o controle dos dados quando eles entram no sistema da empresa.

Para isso é preciso ter mais braços dentro da empresa que devem estar constantemente analisando e qualificando dados. Não significa que só o seu time de TI vai trabalhar nisso, mas sim, grande parte dos colaboradores de diversas áreas.

Mas não é preciso uma forte movimentação para capacitar os funcionários de outras áreas para lidar com dados? A resposta é sim e não ao mesmo tempo: os colaboradores de fora não precisam aprender a programar, mas precisam ter noção da importância dos dados qualificados.

Com uma interface amigável e intuitiva, o framework colaborativo dá a autonomia necessária para que todos os setores aproveitem dos dados e, ao mesmo tempo, trabalhar para que eles sejam cada vez mais qualificados.

É claro que tudo isso é feito de forma segura, orientando o acesso a dados que fazem sentido para os colaboradores que estão se beneficiando deles. Como dissemos anteriormente, um funcionário do financeiro não tem acesso às mesmas informações que as pessoas que trabalham no marketing.

Agora que já entendemos os diferentes tipos de framework, vamos para a prática. Veja a seguir como uma estratégia de governança de dados colaborativa precisa ser implementada.

Mão na massa: o passo a passo da implementação

Entendendo o cenário atual

O alto volume de dados e repositórios é vantajoso por enriquecer, mas por outro lado pode poluir os data lakes. Para esse primeiro momento é preciso entender como os dados brutos estão chegando nos repositórios para aí sim se preparar para o segundo passo.

Um estudo pela IDC mostra que o tempo dos profissionais de dados, em sua grande maioria é composto por:



Mão na massa: o passo a passo da implementação

Isso mostra que muito tempo é desperdiçado, levando em consideração que o estudo estima o tempo gasto em análise, preparação e proteção de 24h semanais. Isso acontece pois, quando as pessoas não encontram dados que precisam, elas recriam em valores e regras que vão ficando cada vez mais distintas entre si.

Com uma boa ferramenta é possível criar essas regras e valores que, ao qualificar dados de forma automática, vão mantendo eles muito mais organizados e dentro do mesmo pipeline da política de dados da empresa.

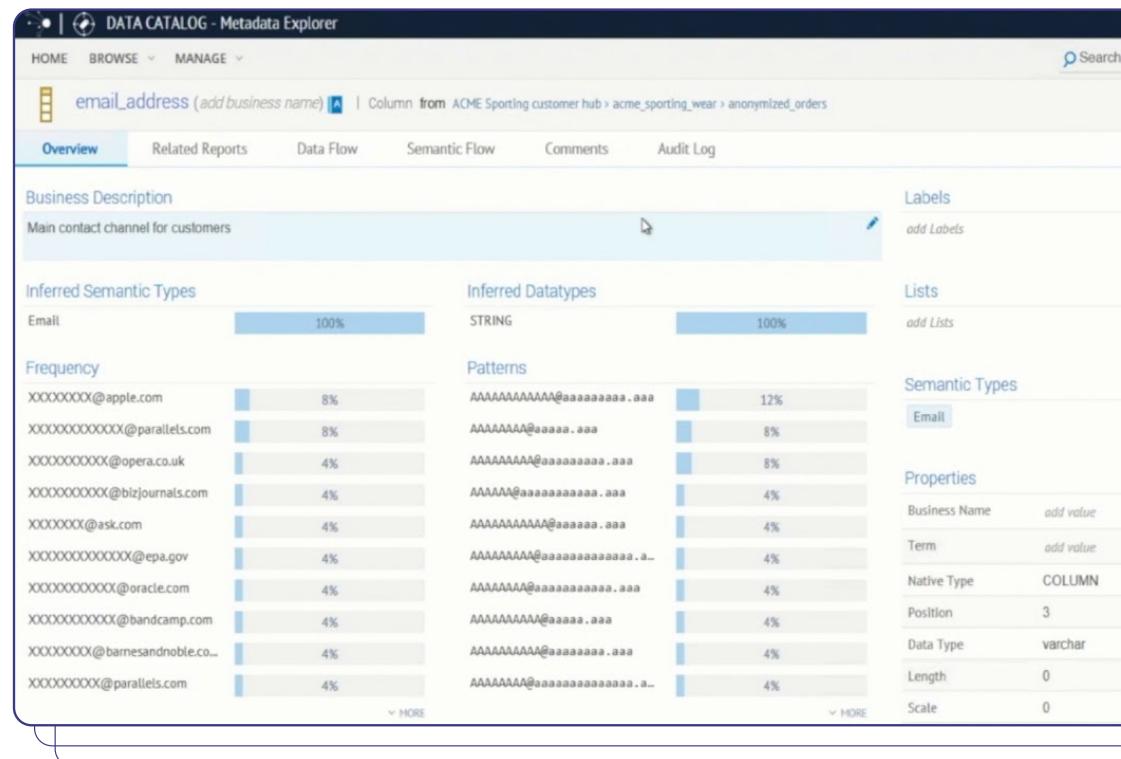
Aplicando o Data Profiling

Para descobrir, entender e organizar os dados é preciso fazer o Data Profiling. Essa prática é um dos primeiros passos para começar a verificar as informações que estão duplicadas nas diversas fontes de dados.

Normalmente esse processo é feito manualmente pelos especialistas de dados, porém, a demanda acaba sendo dificílima de atender. Por isso existem ferramentas de data profiling, como o Talend Data Catalog e Talend Data Preparation, que ajudam na organização de forma inteligente e automatizada para documentar todos as ações partindo dos data lakes até os datasheets.

Além disso, esse tipo de ferramenta ajuda a mostrar os dados de forma muito mais legível para as pessoas de áreas menos técnicas da organização. Com uma boa estratégia e ferramentas confiáveis é possível ganhar um timing de mercado incrível.

Mão na massa: o passo a passo da implementação



Escalando o acesso a dados

Na maioria dos casos, as pessoas que conhecem melhor os dados não são do TI e, consequentemente, não são especialistas de dados. Com um bom framework de data governance é possível empoderá-las e fazer com que elas ajudem no processo.

É possível ter um approach aos dados de forma mais simples ao utilizar interfaces amigáveis para as pessoas de fora do TI. Dessa forma elas conseguem contribuir para a qualificação e aumentar ainda mais a velocidade de processamento e a confiabilidade das informações geradas.

Isso não significa que todo o board da empresa terá acesso a todos os tipos de dados, mas sim que cada um trabalhará com os dados que fazem sentido para sua operação. Essa prática é essencial para manter o acesso inteligente às informações que podem ser sigilosas e evitar problemas futuros com a LGPD.

Nessa etapa os heads de cada setor terão acesso aos dados e ajudarão nessa qualificação. É o conceito de Data Owners que citamos anteriormente. Seguindo a política de dados da empresa, cada head qualifica os parâmetros que são necessários para o desenvolvimento do seu trabalho.

Protegendo e mascarando os dados

Agora que os dados estão sob controle é possível identificar os padrões de organização e deixá-los seguros para que todos os colaboradores tenham acesso às informações. Antes de facilitar o acesso precisamos nos certificar de que temos segurança.

É aí que os dados são mascarados e criptografados para rodar livremente pela companhia. O time de dados delega responsabilidades de proteção de dados que, num primeiro momento, são feitas de forma automática seguindo as guidelines da política de dados.

Mesmo com as automações, os data owners conseguem mascarar de forma simples algumas subcategorias de dados antes de repassar o acesso. Imagine que em uma campanha de marketing para um evento alguns participantes não concordaram em compartilhar seus dados. Essa seleção pode ser feita em poucos cliques pelo gestor de marketing para mascarar esses dados e passá-los de forma segura para que uma third party não tenha acesso.

Facilitando o acesso para todas as áreas da empresa

Com um fluxo de dados bem definido é muito mais fácil liberar o acesso aos dados para os colaboradores. Já que temos dados confiáveis e seguros, é hora de definir pipelines automatizados que liberam o acesso em diferentes níveis e subcategorias para os funcionários.

Dentro de cada área é possível hierarquizar o acesso às informações para os internos e, ao mesmo tempo, separar o tipo de dado que vai para cada área. Por exemplo: todos do setor de inovação em uma companhia têm acesso a tipos específicos de informação sobre os clientes e, dentro desse setor, os estagiários têm um nível de acesso diferente dos diretores.

Tudo isso pode ser pré-definido de maneira simples e ágil, onde um acesso é gerado de acordo com a área e a hierarquia do colaborador. Com esse acesso único para cada um na empresa, fica muito mais fácil controlar onde cada tipo de dado está chegando e também qual é a sensibilidade desses dados.

Com o papel definido dentro do ecossistema do negócio, o controle sobre o que está sendo feito é muito maior e abre pouquíssimas margens para o erro. Lembre-se: democratizar o acesso é muito diferente de liberá-lo sem parâmetros de segurança.

CASE

areeba

Potencializando a operação com o Talend

Veja como uma das maiores empresas financeiras do Oriente Médio acelerou seus processos e fortaleceu o compliance através do Talend Data Catalog.

INDÚSTRIA

Finanças | Setor Bancário

NECESSIDADES

- Gerenciamento de riscos de fraude
- Prevenção à lavagem de dinheiro
- Eficiência de marketing e vendas
- Compliance e leis de proteção de dados

DESAFIO

Construir um data lake sólido e implementar uma governança de dados eficaz

PRODUTOS UTILIZADOS

- Talend Data Integration
- **Talend Data Catalog**
- Talend Data Quality
- Talend Data Stewardship
- Talend Data Preparation

Criando uma cultura que valoriza a governança de dados

areeba é uma empresa focada em melhorar os processos de pagamentos para bancos, comércios, agências do governo e até mesmo pessoas físicas por meio da tecnologia.

Para isso acontecer da melhor forma, Elie Soukayem – diretor de Data Analytics da empresa – entende que é preciso oferecer uma plataforma segura e completa para que eles possam gerenciar seus cartões de crédito.

CASE - areeba | Potencializando a operação com o Talend

[CASE] areeba | Potencializando a operação com o Talend

Construindo um data lake completo para permitir uma boa estratégia de governança

Antes de tudo foi preciso conscientizar todos os colaboradores de que os dados são o grande combustível para ter bons resultados dentro da organização.

Com o apoio do CEO da empresa, Soukayem construiu um time de dados do zero. Na areeba, o time de TI concentra operações e fontes de dados. Abaixo nesse guarda-chuva, o time dados assegura o compliance, ciclos de dados e principalmente a governança.

Na criação do data lake, o Talend foi peça fundamental com suas APIs para construir uma estrutura rápida, intuitiva e que pudesse integrar, limpar e estruturar todos os dados que entram no data lake em tempo recorde.

GOVERNANÇA DE DADOS | **Implementando uma estratégia de sucesso**

Com a ferramenta foi possível centralizar todas as informações em um data lake e, de forma automática garantir que os dados estavam seguros e limpos de acordo com as necessidades da empresa. Confira a seguir os resultados obtidos:

- Ativação de um data lake com 500 GB de capacidade, com uma média de **1 a 2 GB de informação transferidos e ingeridos diariamente.**
- Ingestão de dados a cada 5 ou 10 segundos para evitar fraudes
- Estratégia sólida de governança de dados focada em compliance sem quebrar as leis de proteção de dados

“Como uma empresa de tecnologia e finanças, temos necessidades específicas para data governance, por conta da natureza dos dados que armazenamos. O Talend foi essencial para manter o compliance e nos trazer fôlego para desenvolver novos produtos para os clientes”

Elie Soukayem,
Diretor de Data Analytics na **areeba**

Conclusão

Governança de dados é um ótimo caminho para empresas que buscam atingir resultados fora da curva. O processo pode parecer longo e difícil, mas com os parceiros de negócio certos você impulsiona os rendimentos da organização de uma maneira surpreendente.

Vimos também que a escolha da ferramenta certa é essencial para que corra tudo bem na estratégia. Nós da Sysvision somos representantes oficiais do Talend no Brasil, focados 100% em ajudar empresas que buscam estar entre as melhores com ajuda das melhores ferramentas.



Sysvision International

A Sysvision é uma empresa multinacional focada em soluções de Data Management & Analytics proporcionando aos clientes um diferencial competitivo com informações estratégicas obtidas através de processos tecnológicos.

Foi criada no ano de 2001 em Portugal com a missão de ser uma das principais empresas de desenvolvimento de projetos de Telecom para a Europa. Em 2007, após uma forte expansão em toda a Europa, a Sysvision iniciou suas atividades no Brasil direcionando seus esforços no mercado de Analytics.

Com apoio de parceiros locais e internacionais a empresa cresceu rapidamente no mercado brasileiro tornando-se a principal parceira de empresas como Talend®, AtScale, HCL e Huawei Cloud. Rapidamente a Sysvision expandiu seus serviços para toda a América do Sul.



www.sysvision.com.br