



INTRODUÇÃO AO DATA MESH

*Um guia essencial para a
gestão de dados em 2023*

UMA MUDANÇA DE PARADIGMA _____ 3

O QUE É O DATA MESH? _____ 6

QUANDO IMPLEMENTAR O DATA MESH _____ 7

O PROBLEMA DOS DATA LAKES & DATA WAREHOUSES _____ 9

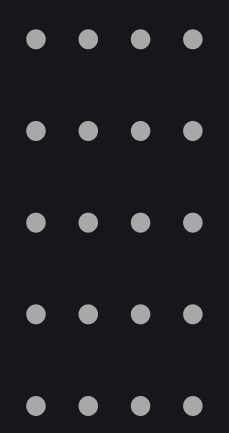
A ATUAL ARQUITETURA DE DADOS MULTIMODAL NA NUVEM _____ 10

OS 4 PRINCÍPIOS DO DATA MESH _____ 11

ARQUITETURA DATA MESH NA PRÁTICA _____ 15

BENEFÍCIOS DO DATA MESH _____ 16

COMO A BRQ PODE AJUDAR _____ 17



Uma mudança de paradigma

O investimento em Data & Analytics e Data Science aumentou drasticamente em todos os setores, mas não renderam o resultado que as empresas esperavam.

Enquanto isso, o volume, variedade e fontes de dados cresceram na mesma proporção. O controle dessas informações continuam mantidos nas mãos de poucas pessoas, geralmente nos departamentos de TI.

A busca pela cultura data-driven leva as unidades de negócio, como marketing, vendas e executivos a consumirem todos os tipos de dados para tomada de decisão, mas seguindo o modelo tradicional, ficam dependentes da TI para centralizar e conectar todas as informações necessárias.

Do lado da TI, um time de engenheiro de dados fica responsável por dar vazão à essas demandas, porém, eles não possuem o conhecimento de negócio necessário para gerar verdadeiros insights, além de sobrecarregar a TI e prolongar a tomada de decisão.

• • • •
• • • •
• • • •
• • • •
• • • •
• • • •
A estrutura centralizada de Data Lake/Warehouses, como é tida hoje, trouxe desafios para a democratização da informação:

- Falta de controles eficazes para democratização e manejo do dado pela área de origem;
- Problemas de qualidade dos dados, uma vez que o time de engenharia de dados não conhece os dados como a área que a gerou;
- Escalabilidade do dado à nível organizacional.



A resposta tradicional a esses problemas de escala e complexidade tem sido voltar nossa atenção para plataformas mais robustas e supostamente projetadas para aliviar esses problemas.

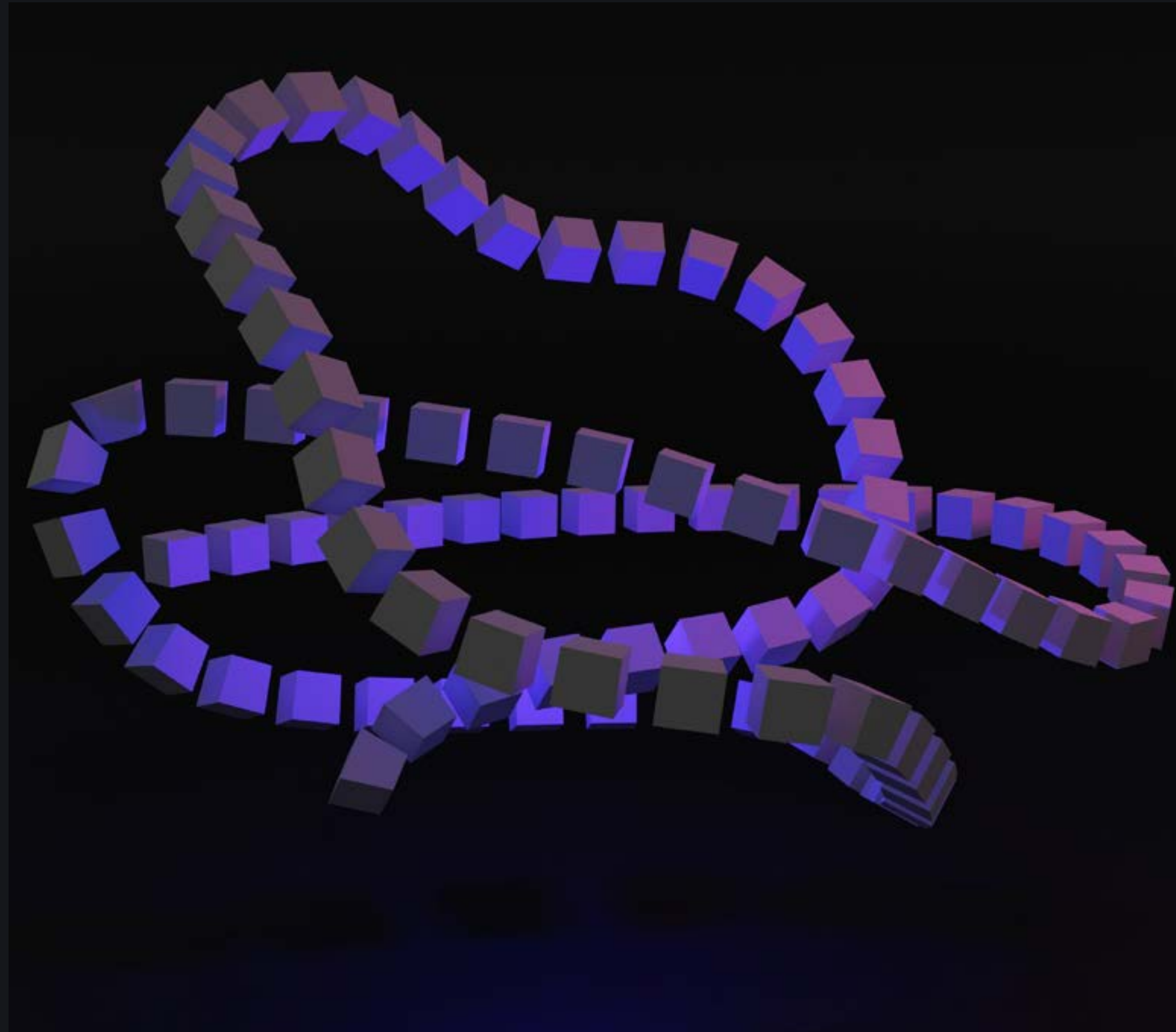
Porém, infelizmente, essas novas soluções geralmente introduzem novos problemas, não aproximando as organizações de seu objetivo final.

“Você acaba nessa armadilha contínua de tentar inicializar sua próxima plataforma de big data, nunca se materializando em arquitetura e nunca realmente vendo valor no final”, diz Zhamak Dehghani, diretora de tecnologias emergentes da ThoughtWorks North America.

Essa espiral de ineficiência e insatisfação entre grandes empresas deu origem ao Data Mesh, uma abordagem nova e mais holística para gerenciar escala em empresas globais.

O conceito Data Mesh é uma mudança de paradigma que exige mudanças organizacionais, tecnológicas e de arquitetura. No entanto, a lógica inerente ao Data Mesh, combinada com o sucesso de várias empresas que o colocaram em prática, sugere





que ele tem o potencial de ajudar as organizações a alcançar o tipo de resultado que esperam desde o início da era do Big Data.

Enquanto as plataformas de dados atuais são amplamente centralizadas e monolíticas, construídas em torno de pipelines complexos que alimentam dados em um armazenamento central sem levar em conta a qualidade e a eficiência, o Data Mesh incentiva o cultivo de diferentes conjuntos de dados como produtos distribuídos orientados por domínios. Cada um desses conjuntos de dados específicos de domínio tem seus próprios engenheiros incorporados e proprietários de produtos que são responsáveis por gerenciar esses dados e sua disponibilidade para outras equipes. Ele se baseia em um nível de propriedade e responsabilidade de dados ausente em abordagens anteriores.

O que é o Data Mesh?

Data Mesh, criado por Zhamak Dehghani, e também conhecido em português como Malha de Dados, é um tipo de arquitetura voltada para gestão corporativa descentralizada. Nele, o caminho entre stakeholders e a informação de que necessitam para realizar tarefas e tomar decisões é o menor possível.

O objetivo da estrutura de Mesh é tornar o banco de dados mais horizontal e democrático. Utilizando soluções modernas de TI, como controle de acesso e automatização de processos, é possível nivelar a prioridade de acesso para todos os colaboradores da empresa sob demanda.

O Data Mesh é uma evolução do Big Data consolidado nas empresas. O negócio tem o mesmo poder de coleta e Data Analytics em grandes volumes de dados, mas nesse caso, mais bem distribuídos. Assim, o poder de interpretação e reação de cada profissional conectado ao sistema se torna mais independente e efetivo dentro de insights segmentados e relacionados diretamente às suas funções.

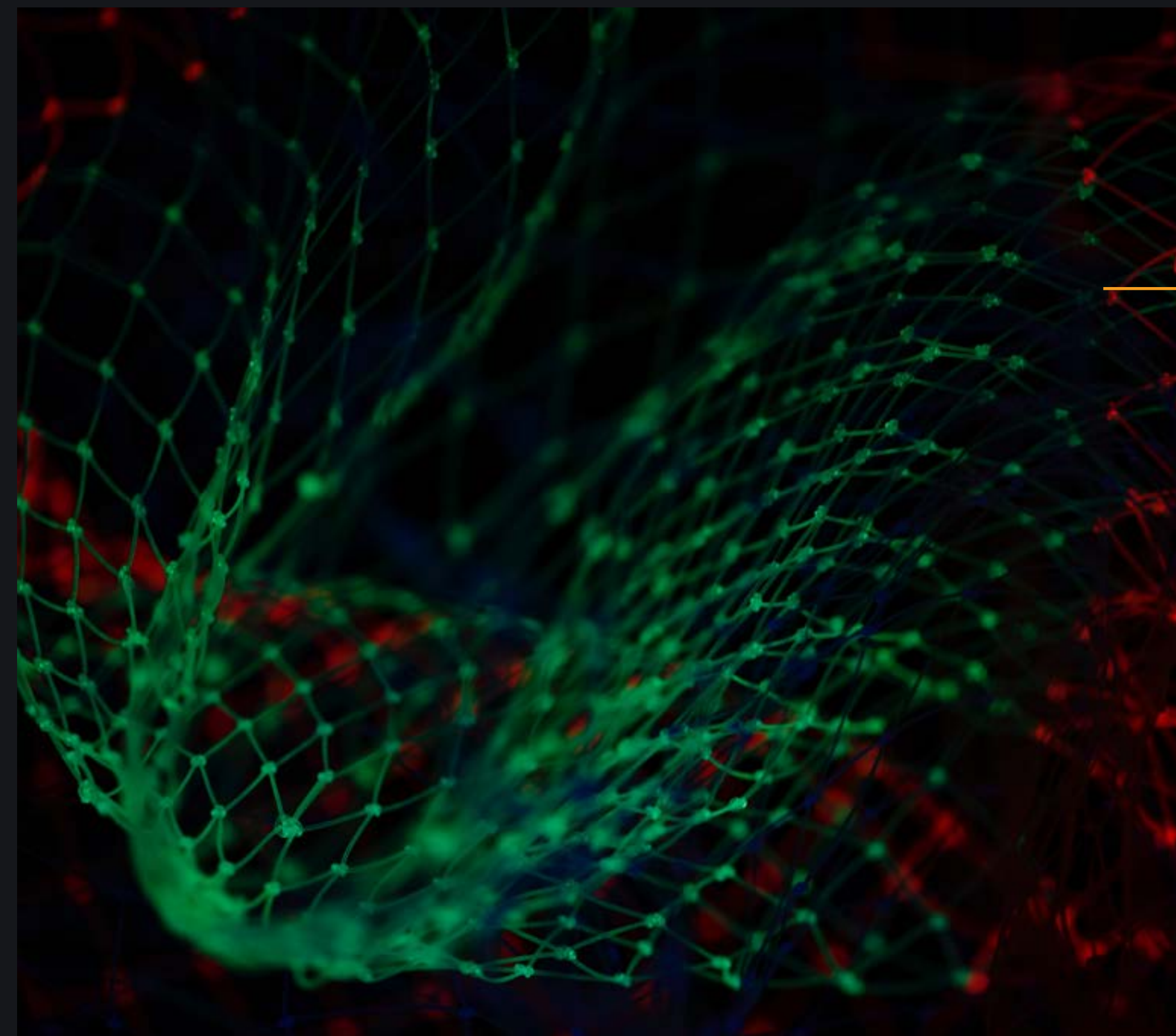


Quando implementar o Data Mesh



O Data Mesh é ideal para organizações que lidam com:

- Mudança constante na topologia do seu cenário de dados;
- Proliferação de fontes de dados e consumidores;
- Diversidade nas necessidades de transformação e processamento de dados;
- A necessidade de responder rapidamente as mudanças relacionadas a dados.





Se você tiver mudanças contínuas em seu cenário de dados, juntamente com uma proliferação de fontes, e estiver insatisfeito com o investimento em Data Science em relação aos resultados alcançados, vale a pena considerar a adoção de uma abordagem Data Mesh.

Caso o seu negócio tenha um desenvolvimento orientado ao domínio, começou a trabalhar com microsserviços ou se fez uma migração para a nuvem, a migração para o Data Mesh tende a ter um custo bem menor.

De modo geral, o Data Mesh exige uma mudança cultural dentro da sua empresa, além de uma nova arquitetura e infraestrutura, por isso deve ser trabalhado com cautela.

O problema dos Data Lakes & Data Warehouses



As deficiências do Data Warehouse acabaram gerando o conceito de Data Lake.

A ideia era extrair os dados dos sistemas operacionais, carregá-los em formato semiestruturado e bruto, e depois deixar para outros sistemas fazerem qualquer limpeza, transformação ou criar data warehouses para acesso SQL. A desvantagem dessa abordagem é que a responsabilidade pelos dados fica ofuscada. À medida que a quantidade de dados e o tamanho de uma empresa aumentam, os diferentes grupos dentro da empresa perdem a noção de quem possui e cuida dos dados.

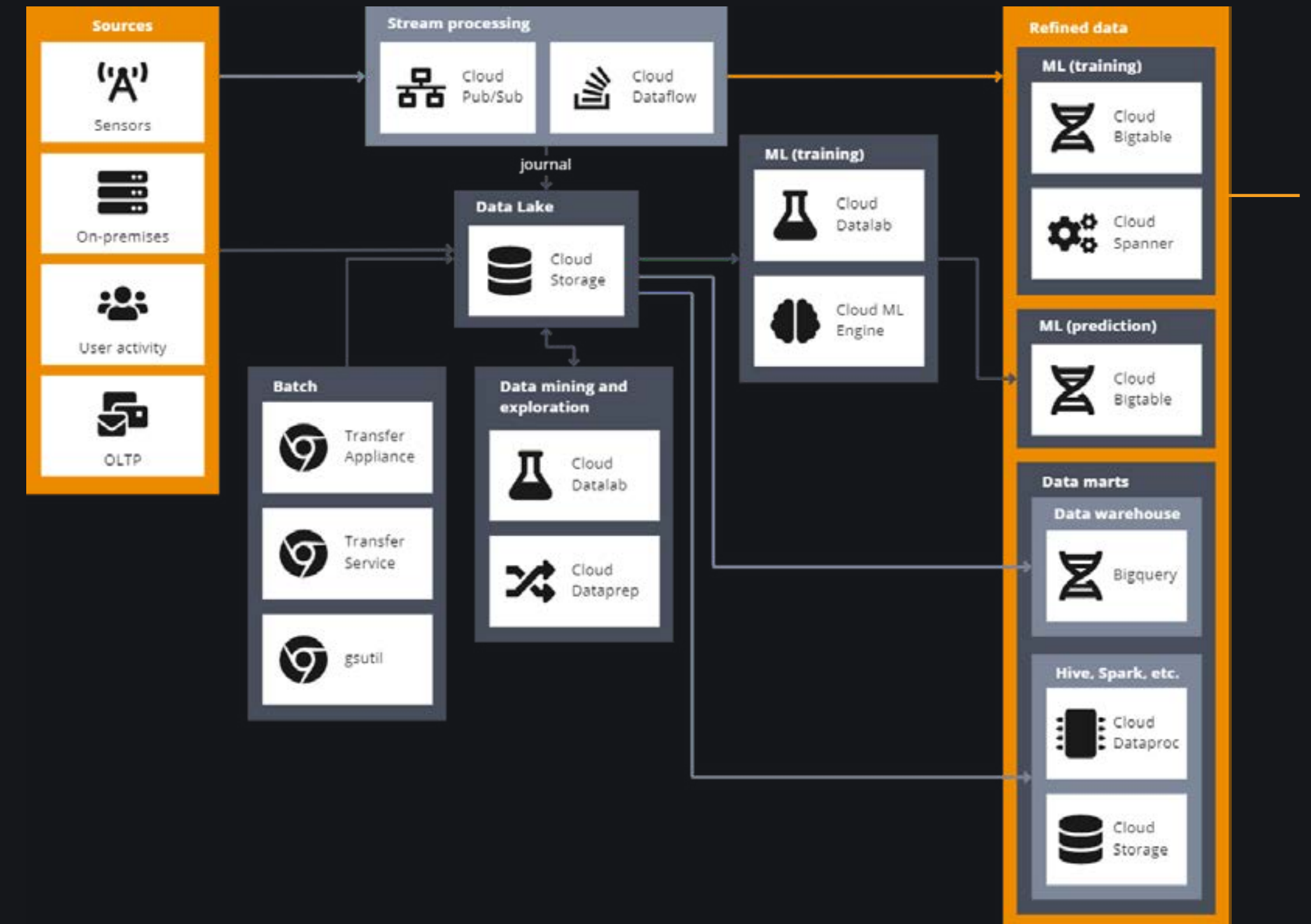
Todos continuam a operar como se os dados chegassem da forma correta. Porém, como ninguém se apropriou desses conjuntos de dados, a integridade da informação pode sofrer impactos negativos. O que temos é o gerenciamento de dados de passar a responsabilidade.

A diferença entre Data Mesh e outras arquiteturas

Quando analisamos a infraestrutura de coleta e armazenamento de dados, não há muita diferença entre Data Mesh e outras arquiteturas populares, sob o ponto de vista de Data Science. Investir em um sistema robusto de informação é fundamental para qualquer empresa que deseja ser competitiva no futuro.

A grande diferença está na outra ponta: na maneira como essas informações são utilizadas dentro da rotina do negócio. Uma das maiores evoluções do Data Mesh é sua organização em relação ao já conhecido Data Lake. A forma como esses dados são tratados ao serem centralizados em um sistema de informação é o que mais o diferencia.

O Data Lake, apesar de permitir a interação entre todos os tipos de dados, não tem o mesmo pré-processamento da malha de dados, fazendo com que essas informações, muitas vezes, percam seu contexto. Como dissemos, Data Mesh apresenta o conceito de propriedade, ou seja, cada departamento é responsável por dar contexto e significância para seus ativos.



Fonte: <https://cloud.google.com/architecture/build-a-data-lake-on-gcp>



Já em relação aos silos, ou Data Warehouse, o Mesh acaba com as barreiras entre diferentes camadas de credenciais e departamentos. Essa separação pode até ser interessante para focar a gestão de dados, mas não é mais tão eficiente para o mercado moderno.

Afinal, uma empresa digitalizada não tem mais ilhas operacionais. Tudo o que um departamento faz se reflete nos outros, como um grande organismo único. É por isso que o Data Mesh vem ganhando tanta relevância nos últimos anos.

Os 4 princípios do Data Mesh

Essa abordagem do Data Mesh é baseada em 4 princípios:

1 - ARQUITETURA DE DADOS ORIENTADA AO DOMÍNIO

O Data Mesh exige uma mudança de conjuntos de dados centralizados para conjuntos de dados orientados ao domínio. Em vez de canalizar dados de vários domínios para uma plataforma monolítica e afastá-los dos proprietários originais, os especialistas do domínio que conhecem melhor os dados permanecem no controle de seus dados.

Cada domínio pode compartilhar seus dados através de APIs ou end-points, por exemplo, para serem utilizados em várias áreas da empresa. Isso implica que a arquitetura deve remover qualquer acoplamento para permitir que os domínios forneçam seus dados analíticos e permita que a engenharia concentre em gerenciar os dados, independentemente de outros domínios. Para escalar, a arquitetura deve oferecer suporte à autonomia das equipes de domínio em relação à implantação de seus sistemas de dados operacionais ou analíticos.

2 - DADOS DISPONIBILIZADOS COMO UM PRODUTO

A criação de conjuntos de dados orientados ao domínio não elimina a necessidade de limpar, preparar, agregar e servir dados. A diferença é que esses processos são gerenciados internamente, dentro de cada domínio.

Um temor dessa abordagem orientada ao domínio é que ela leve a novos silos de dados à medida que diferentes grupos adotam diferentes soluções de tecnologia para atingir esses objetivos internos. O segundo princípio do Data Mesh, que exige uma mudança de pensamento em relação aos dados como um produto, combate esse efeito de silo. Quando uma organização considera seus conjuntos de dados como produtos que precisam ser valiosos para todos os usuários internos, de analistas a cientistas de dados, qualquer efeito do tipo silo seria reconhecido como uma falha – um produto que ninguém pode acessar ou usar.

Esses conjuntos de dados orientados ao produto devem ser:

DETECTÁVEL

Os proprietários de domínio precisam garantir que seu produto seja fácil de encontrar, por meio de um registro, um catálogo de dados ou de alguma outra forma.

AUTODESCRITIVO

Engenheiros de dados e cientistas de dados devem ser capazes de descobrir, entender e consumir conjuntos de dados com facilidade.

ENDEREÇÁVEL

Convenções de nomenclatura comuns devem ser usadas para tornar o produto de dados acessível programaticamente, especialmente porque domínios diferentes podem adotar formatos de dados diferentes.

INTEROPERÁVEL

Os usuários devem ser capazes de correlacionar dados entre domínios e adicionar, filtrar, agregar, etc.; aplicar padrões de governança global para interoperabilidade e comunicação é essencial.

CONFIÁVEL

Os produtos de dados precisam ser verdadeiros em termos de precisão dos dados, com objetivos de nível de serviço definidos e monitorados.

SEGURO

A política de controle de acesso global por meio de SSO e o controle de acesso baseado em função garantirão o acesso seguro aos conjuntos de dados.

Essa mudança de pensamento em direção a conjuntos de dados orientados ao domínio como produtos exige mudanças organizacionais, incluindo novas funções, como proprietários de produtos de dados e desenvolvedores de produtos de dados. Ele inverte o modelo de servir dados de orientados à tecnologia para orientados ao domínio, e isso significa que os proprietários de produtos de dados precisam fornecer os mesmos dados em diferentes modos de acesso para satisfazer as ferramentas nativas de cientistas de dados, analistas de dados ou outros usuários finais.

3 - PLATAFORMA DE DADOS SELF-SERVICE

A distribuição de propriedade e arquitetura requer um novo tipo de infraestrutura que se afasta do paradigma de obter tudo em um só lugar para uma unidade de arquitetura que pode incluir dados, o código que os processa e os atende e a infraestrutura que o dirige e o protege.

Pedir a cada grupo de domínio para construir e projetar essa infraestrutura seria ineficiente. Em vez disso, uma equipe central de infraestrutura de dados deve possuir e fornecer a tecnologia que os domínios precisam para seus produtos de dados.

A infraestrutura precisa ser de autoatendimento e generalizada, para que não inclua nenhuma característica específica do domínio que possa torná-la menos utilizável para outro domínio. A lista de recursos é extensa, mas, em última análise, essa infraestrutura de autoatendimento deve facilitar para um grupo criar um novo produto de dados e disponibilizá-lo para as equipes de ciência de dados e BI da empresa.

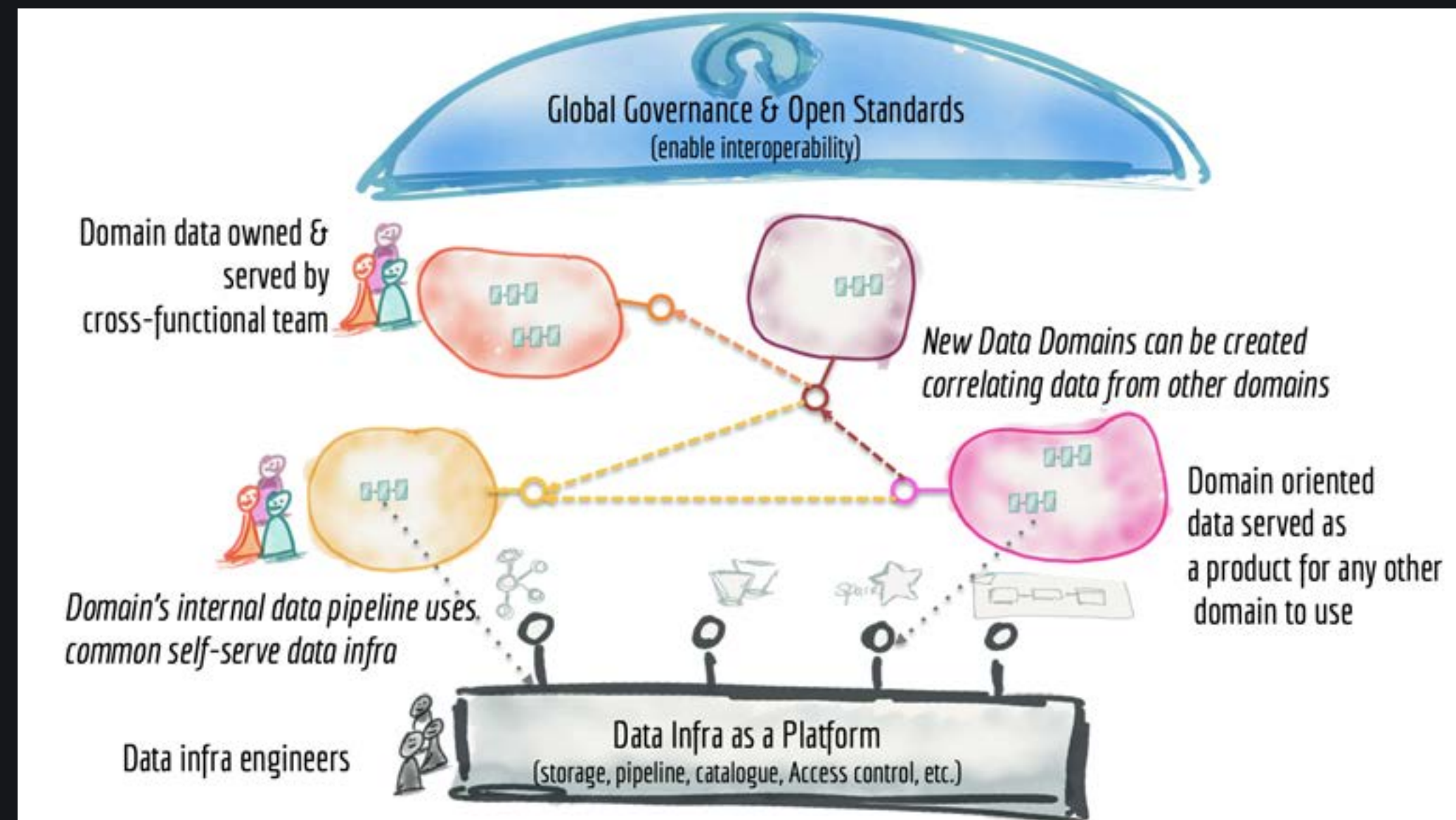


4 - GOVERNANÇA FEDERADA

Como uma organização controla os dados nesse novo ambiente? O Data Mesh implementa um modelo federado no qual os proprietários de produtos de dados concordam com incentivos e regras aplicáveis local e globalmente para garantir a interoperabilidade de seus produtos de dados e outros, juntamente com políticas que governam questões sobre controle de acesso, gerenciamento de esquema e muito mais. Assim, embora os conjuntos de dados existam como domínios independentes, eles são unificados por políticas acordadas.



Como funciona o Data Mesh na prática



Fonte: <https://martinfowler.com/articles/data-monolith-to-mesh.html>

A estruturação de um Data Mesh se inicia como qualquer uma das arquiteturas de dados modernas. Com utilização de soluções de armazenamento, como servidores dedicados ou Cloud Computing, a empresa passa a coletar e armazenar o máximo de informação disponível sobre produção interna, People Analytics, concorrentes e mercado como um todo.

Essa infraestrutura se diferencia a partir desse ponto. Em vez de contar com uma supervisão centralizada e camadas rígidas de acesso, o Data Mesh permite a qualquer colaborador buscar, visualizar e utilizar a informação que precisa para tomar decisões relevantes de negócio.

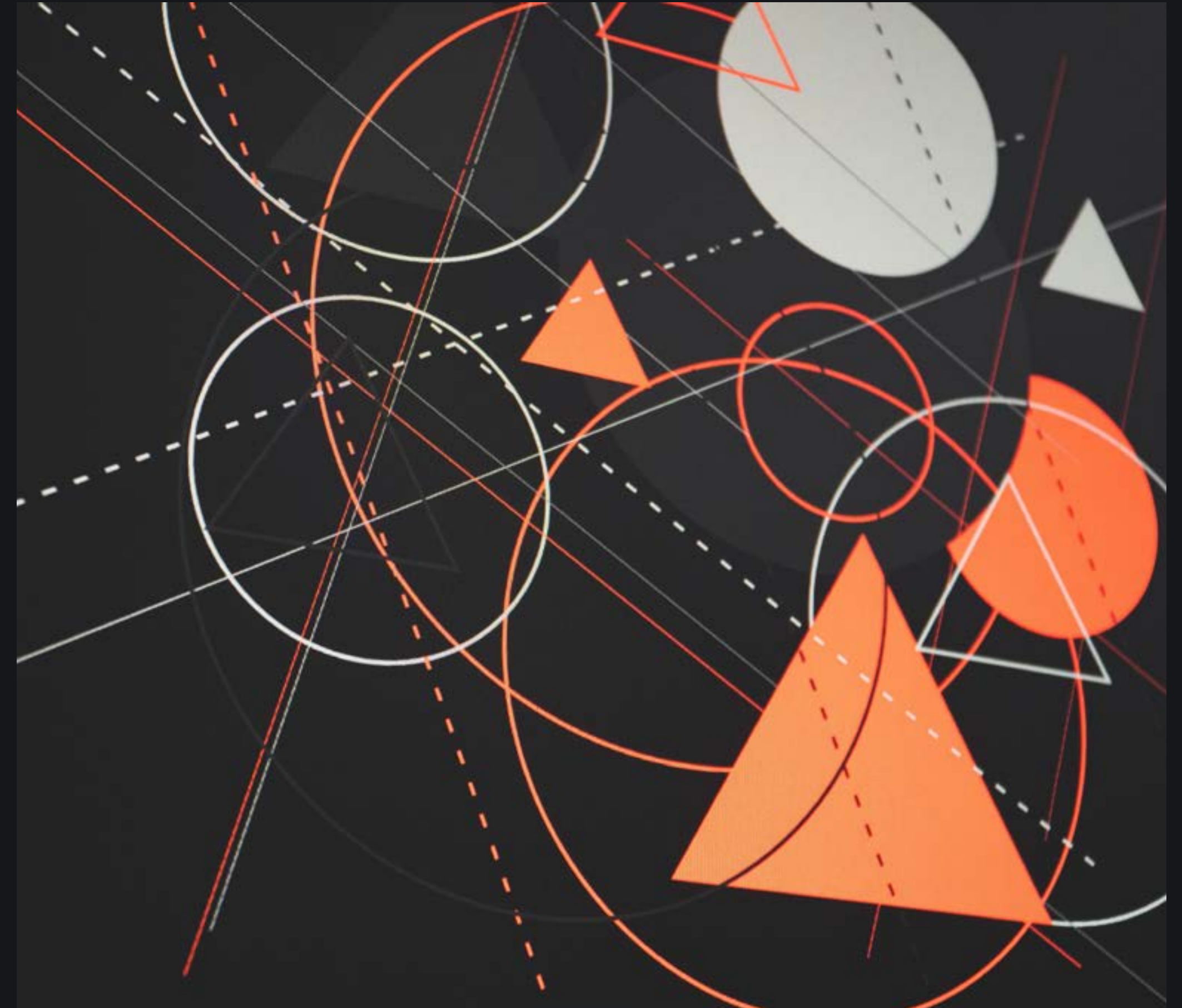
Por exemplo, se um profissional de vendas tem dificuldade nas conversões, pode desconfiar que algo na jornada do cliente criada pelo Marketing está servindo como gargalo. Para identificar o problema, ele pode ir diretamente nas informações coletadas pelo outro departamento, sem a necessidade de solicitar e esperar uma resposta.

Como funciona o Data Mesh na prática

Essa troca constante e direta de informações agiliza processos produtivos e aumenta a proatividade. Mas vale lembrar que, mesmo descentralizado, o Data Mesh tem suas responsabilidades.

Nesse caso que demos como exemplo, o Marketing é responsável por coletar e organizar as informações relacionadas a seu trabalho. Da mesma forma, essas interações com o setor de Vendas devem ser registradas para que possam enriquecer interpretações de todos no futuro.

Com o cuidado com os dados na entrada e a liberdade de utilização na saída, o negócio se torna mais dinâmico e pode acompanhar com mais velocidade as tendências do mercado moderno.



Benefícios do Data Mesh

A implementação eficiente do Data Mesh oferece muitos benefícios:

ACELERA O TEMPO ATÉ O INSIGHT

Há uma redução drástica dos prazos de entrega e permite que os domínios de negócios priorizem e tomem decisões rapidamente. O Data Mesh torna os dados disponíveis para toda a organização e, ao mesmo tempo, oferece liberdade no nível tecnológico para os domínios.

CULTURA DATA-DRIVEN

Se uma empresa deseja se tornar verdadeiramente Data-Driven, o Data Mesh não só aumenta drasticamente a eficiência, reduzindo os custos operacionais e o tempo dos colaboradores, mas também fornece insights de inteligência de negócios muito mais profundos.

DEMOCRATIZAÇÃO DOS DADOS

Se você tiver um grande número de fontes de dados, o processo de dados pode ser altamente complicado. Para produtos de propriedade de dados baseados em domínio que foram federados, o Data Mesh ajuda a automatizar as estratégias certas para torná-las o mais eficiente possível. Assim, o Data Mesh é um passo essencial para melhorar a democratização de dados cruciais.

VISIBILIDADE DE DADOS

Uma dúvida muito comum em quem começa sua pesquisa sobre Data Mesh é se essa liberdade de acesso não torna o sistema inseguro. Na verdade, o que acontece é o contrário. Como todos os colaboradores podem acessar todas as informações, isso significa também que todos podem visualizar e analisar as atividades relacionadas a elas.

Ou seja, qualquer anomalia é rapidamente identificada, e grande parte dos conflitos de registros deixam de existir — já que não se cria um cenário em que mais de um departamento está trabalhando com o mesmo indicador de maneira independente. O Data Mesh, portanto, é bastante seguro.

O principal benefício do Data Mesh não é principalmente técnico, mas organizacional e cultural.



Como a BRQ pode ajudar

Há mais de 29 anos, a BRQ apoia seus clientes na jornada de transformação digital, com soluções end-to-end de cloud, analytics, squads e muito mais.

Estamos à frente das maiores inovações e tendências do mercado, como é o caso do Data Mesh. Possuímos a expertise necessária para apoiar o seu negócio na correta adoção dessa nova arquitetura, do assessment à governança.

Faça como mais de 150 empresas em oito países: fale com nossos #feras e comece a sua jornada de transformação.

[Entre em contato.](#)



Gostou deste material e quer saber mais sobre Data Mesh?

Fale com um especialista
em Data Mesh

ENTRAR EM CONTATO

Referências

1. Dehghani Zhamak, "Data Mesh: Delivering Data-Driven Value at Scale"
2. <https://blog.brq.com/data-mesh/>
3. XConf Brasil 2020, "Data Mesh", <https://youtu.be/3qoG8KR1nwc>
4. <https://starburst.io/resources/data-mesh-whitepaper/>
5. <https://thoughtworks.com/pt/ebook/data-mesh>
6. <https://martinfowler.com/articles/data-monolith-to-mesh.html>
7. <https://cloud.google.com/architecture/build-a-data-lake-on-gcp>