

Características de Dados como Serviço

Úteis Coletar dados de alta qualidade e transformá-los em produtos de dados é uma das melhores maneiras de monetizar dados. No entanto, os produtos de dados são valiosos apenas quando são consumidos para melhorar o desempenho dos negócios. Os produtos de dados são usados? Eles são autônomos ou precisam ser combinados com outros produtos para serem valiosos?

Descobríveis Idealmente, produtos de dados devem ser publicados em um catálogo ou registro onde seja possível pesquisar e explorar. Para permitir que os usuários finais encontrem o que precisam, os produtos de dados devem ser publicados com informações adicionais, como domínio, proprietário, linhagem e métricas de qualidade. Agora, os usuários podem facilmente investigar e encontrar os produtos de dados necessários. **Endereçáveis** A endereçabilidade dos dados está relacionada à nomeação e formatos padronizados. De fato, um produto de dados fornece um endereço permanente e único para o usuário acessar automaticamente, e esse endereço único deve seguir padrões específicos que permitam aos usuários consistentemente acessar todos os produtos de dados, tornando-os endereçáveis.

Compreensíveis Uma vez descoberto um produto de dados, o próximo passo é entendê-lo. Os produtos de dados devem ser documentados e o esquema (representação subjacente dos dados) deve ser descrito. De fato, esquemas de dados com semântica e sintaxe bem descritas permitirão produtos de dados autoatendidos.

Confiáveis Embora a descoberta e a compreensibilidade reduzam a lacuna entre o que o usuário sabe e não sabe sobre os dados, é necessário muito mais para que os dados sejam confiáveis. Uma maneira de fechar essa lacuna de confiança é aderir aos objetivos de nível de serviço de produtos de dados aprovados: Intervalo de mudança e pontualidade Completude Atualidade, disponibilidade geral e desempenho Linhagem

Nativos A usabilidade de um produto de dados está diretamente relacionada à facilidade com que os usuários acessam com suas ferramentas nativas. Esta propriedade refere-se à possibilidade de acessar dados de uma maneira que esteja alinhada com as habilidades e linguagem das equipes de domínio. Por exemplo, analistas de dados provavelmente usarão SQL para construir relatórios e painéis. Cientistas de dados, por sua vez, esperam que os dados venham em uma estrutura baseada em arquivo para treinar modelos de inteligência artificial.

Interoperáveis Os produtos de dados podem ser facilmente combinados? Os metadados são padronizados? Os tipos são padronizados? A implementação de padrões globais harmonizará os dados nos diferentes domínios e estabelecerá interoperabilidade de dados em toda a empresa.

Seguros Os produtos de dados devem ser inerentemente seguros e isso inclui controle de acesso, propriedade e padrões robustos de governança. O acesso pode ser gerenciado centralmente e, em seguida, produtos de dados de domínio podem autorizar o acesso em um nível granular para atender às necessidades específicas das diferentes equipes. Quem pode acessar os dados? Em qual contexto? Qual é o período de retenção? Qual é o nível de confidencialidade?

Descentralização da Propriedade de Domínio dos Dados Uma mudança significativa das arquiteturas de dados tradicionais é que o data mesh promove a propriedade descentralizada de domínio dos dados. “Propriedade” é a palavra-chave, implicando total responsabilidade. O data mesh aplica design orientado a domínio para essa propriedade descentralizada. Cada domínio pode abranger diferentes aspectos, seja de entidade de negócios como vendas ou domínios técnicos como dados de cliques de uma ferramenta de análise. O ponto é estruturar a propriedade com base em domínios de negócios.

Dados como um Produto A descentralização da propriedade de dados já foi tentada, mas frequentemente as equipes geradoras de dados não são incentivadas a assumir essa responsabilidade. O data mesh aplica princípios de pensamento de produto aos dados, tratando-os como um produto em si. Um exemplo é o dado de vendas. Aplicar o pensamento de produto implica entender os usuários, aprimorar com base no feedback e entregar o que é realmente necessário. Uma equipe multifuncional é essencial para essa abordagem.

Infraestrutura de Dados Autoatendida como Plataforma Habilitar equipes de domínio a assumir a propriedade completa dos produtos de dados não é simples. Uma solução é fornecer a

infraestrutura como uma plataforma autoatendida, assim como os provedores de nuvem têm feito. Essa plataforma deve ser agnóstica ao domínio para evitar gargalos e promover escalabilidade.

Governança de Dados Computacional Federada Ao invés da governança de dados centralizada, que frequentemente falha, o data mesh propõe uma abordagem federada. Algumas regras globais são necessárias, mas a ideia é que essas regras sejam mínimas e computacionalmente aplicadas. A governança global deve focar em interoperabilidade, enquanto as questões que podem ser resolvidas dentro de um domínio devem permanecer lá. A verificação da aderência às regras deve ser automatizada. Esse tipo de governança será uma entidade decisória e não uma função centralizada de verificação.

Em resumo, o data mesh visa reestruturar a forma como os dados são tratados nas organizações, descentralizando a propriedade dos dados, tratando os dados como produtos individuais, fornecendo infraestrutura autoatendida e estabelecendo governança federada.