



Big Data Fundamentos 3.0

Big Data Fundamentos – Versão 3.0

O Que é Data Mesh?

Conforme definido pela primeira vez por Zhamak Dehghani, consultora da ThoughtWorks e a arquiteta original do termo, Data Mesh é um tipo de arquitetura de plataforma de dados que abrange a onipresença dos dados na empresa, permitindo um design orientado ao domínio e de autoatendimento. Data Mesh é amplamente considerado a próxima grande mudança arquitetônica em dados.

Data Mesh é uma nova abordagem para projetar e desenvolver arquiteturas de dados. Ao contrário de uma arquitetura centralizada e monolítica baseada em um Data Warehouse e/ou um Data Lake, Data Mesh é uma arquitetura de dados altamente descentralizada.

Data Mesh tenta resolver três desafios quando temos um Data Lake/Warehouse centralizado:

- Falta de propriedade. Quem é o proprietário dos dados - a equipe da fonte de dados ou a equipe de infraestrutura?
- Falta de qualidade. A equipe de infraestrutura é responsável pela qualidade, mas não conhece bem os dados.
- Escalonamento organizacional. O armazenamento central torna-se o gargalo, como no caso de um Data Lake/Warehouse empresarial.

O objetivo com Data Mesh é tratar os dados como um produto, com cada fonte tendo seu próprio gerente/proprietário de produto de dados (que fazem parte de uma equipe multifuncional de Engenheiros de Dados) e sendo seu próprio domínio claramente focado e com uma oferta autônoma, tornando-se os blocos de construção fundamentais de uma malha (Mesh), levando a uma arquitetura distribuída orientada por domínio.

Observe que, por motivos de desempenho, você pode ter um domínio que agrupa dados de várias fontes. Cada domínio deve ser detectável, endereçável, autoexplicativo, seguro (governado por controle de acesso global), confiável e interoperável (governado por um padrão aberto). Cada domínio armazenará seus dados em um Data Lake e, em muitos casos, também terá uma cópia de alguns dos dados em um banco de dados relacional.

Outro componente do Data Mesh é a infraestrutura de dados como plataforma, que fornece armazenamento, pipeline, catálogo de dados e controle de acesso aos domínios. A ideia principal é evitar a duplicação de esforços. Isso permitirá que cada equipe de produto de dados crie seus produtos de dados rapidamente.

Vale ressaltar que Data Mesh ainda é uma tendência e sua implementação tem diversos desafios técnicos. Mas toda tecnologia de sucesso um dia foi uma tendência. Como dizia Bill Gates, fundador da Microsoft: O segredo do sucesso em tecnologia está em reconhecer o potencial de uma tendência no futuro e chegar lá primeiro.