

Data Hub

De acordo com o Goasduff (2020), *Data Hubs* são “hubs” conceituais, lógicos e físicos para mediação da semântica, em suporte à governança e compartilhamento de dados entre sistemas. *Data Hubs* possibilitam o fluxo contínuo de dados devidamente governados.

Para o DAMA (2017), existem diferentes tipos de *Data Hubs*, uma vez que este é um conceito que objetiva a consolidação e o provimento de dados consistentes tanto para outros sistemas quanto diretamente para humanos. Neste sentido, DAMA (2017) menciona que *Data Warehouses* e soluções de Gestão de Dados Mestres são tipos de *Data Hubs*.

Segundo Harrad (2020), o Data Hub é o local ideal para os dados principais de uma organização. Ele centraliza os dados que são críticos entre os aplicativos e permite o compartilhamento contínuo entre diversos *endpoints*, ao mesmo tempo em que é a principal fonte de dados confiáveis para a iniciativa de governança de dados. Os *Data Hubs* fornecem Dados Mestres para aplicativos e processos corporativos. Eles também são usados para conectar aplicativos de negócios a estruturas analíticas, como *data warehouses* e *data lakes*.

Figura 1: Data Hub como pilar principal de dados governados. Fonte: HADDAD (2020).



	Data Hub	Data Warehouse	Data Lake
Uso primário	Processos operacionais	Análise e relatórios	Análise, relatórios e <i>Machine Learning</i> .
Formato dos dados	Estruturados	Estruturados	Estruturados e não estruturados
Governança de Dados	Pilar principal para todas as regras de aplicação de governança de dados	Governança pós-fato, pois consome dados operacionais existentes	Abordagem de dados “Use por sua conta e risco”. Pouco governado
Qualidade de Dados	Altíssima qualidade	Alta qualidade	Média e baixa qualidade

<p>Integração com aplicações</p>	<p>Integração bidirecional em tempo real com processos de negócios existentes por meio de APIs.</p>	<p>ETL monodirecional em lote. Os dados transformados e limpos são atualizados em baixa frequência (por hora, diariamente ou semanalmente)</p>	<p>ETL ou ELT monodirecional em lote. Os dados são despejados sem controle, assumindo uma limpeza futura pelo consumidor</p>
<p>Interação com usuários de negócio</p>	<p>Pode ser a fonte primária de autoria de elementos-chave de dados, como Dados Mestre e Dados de Referência. Expõe interfaces amigáveis para criação de dados, administração de dados e pesquisa.</p>	<p>Oferece acesso somente leitura a dados limpos e preparados, por meio de relatórios, painéis analíticos ou consultas ad-hoc.</p>	<p>Requer limpeza/preparação de dados antes do consumo. O acesso aos usuários corporativos é oferecido principalmente por meio de relatórios, painéis ou consultas ad-hoc. Usado para preparar conjuntos de dados de aprendizado de máquina.</p>
<p>Processos Operacionais da organização</p>	<p>Repositório principal para dados confiáveis expostos em processos de negócios. Pode ser o principal condutor dos processos de negócios corporativos.</p>	<p>Serve principalmente para processos de análise.</p>	<p>Atende principalmente processos de <i>Machine Learning</i>.</p>

Revisão #8

Criado Mon, Mar 21, 2022 10:19 AM por FLAVIO LOPES DE MORAIS

Atualizado Wed, Dec 14, 2022 9:02 AM por FLAVIO LOPES DE MORAIS