

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*.

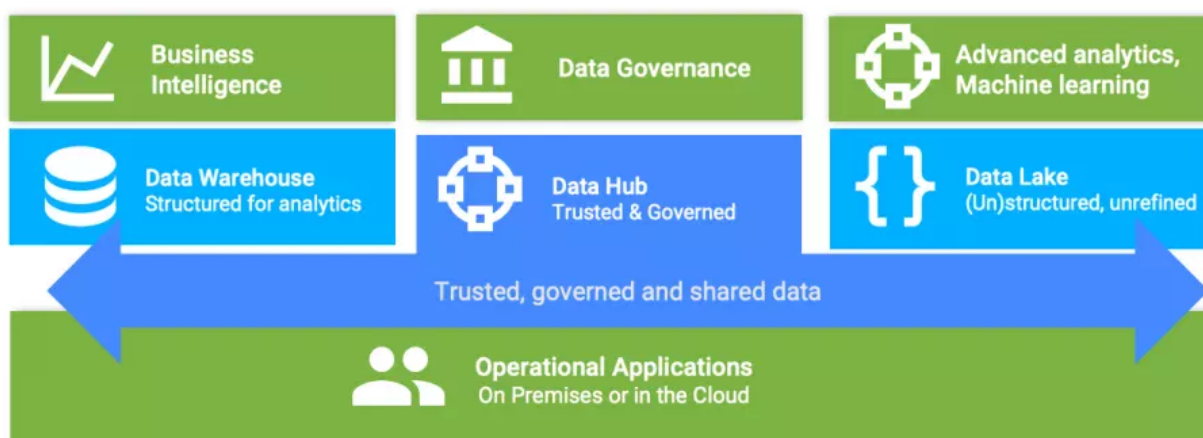


Figura 1: *Data Hub* como pilar principal de dados governados. Fonte: HADDAD (2020).

Diferentemente de *Data Warehouses* e *Data Lakes*, que são focados no provimento de dados para finalidades analíticas, os *Data Hubs* servem como pontos de mediação e compartilhamento de dados, com o foco em governança (HARRAD, 2020).

Data warehouses, *Data Lakes* e *Data Hubs* não são alternativas intercambiáveis, mas sim complementares. Juntos podem apoiar iniciativas baseadas em dados e transformação digital. A tabela abaixo resume suas semelhanças e diferenças HARRAD (2020):

	<i>Data Hub</i>	<i>Data Warehouse</i>	<i>Data Lake</i>
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
Integração com aplicações	Integração bidirecional em tempo real com processos de negócios existentes por meio de APIs.	ETL monodirecional em lote. Os dados transformados e limpos são atualizados em baixa frequência (por hora, diariamente ou semanalmente)	ETL ou ELT monodirecional em lote. Os dados são despejados sem controle, assumindo uma limpeza futura pelo consumidor
Interação com usuários de negócio	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.	Oferece acesso somente leitura a dados limpos e preparados, por meio de relatórios, painéis analíticos ou consultas ad-hoc.	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.
Processos Operacionais da organização	Repositório principal para dados confiáveis expostos em processos de negócios. Pode ser o principal condutor dos processos de negócios corporativos.	Serve principalmente para processos de análise.	Atende principalmente processos de <i>Machine Learning</i> .

Revisão #8

Criado 2022-03-21 10:19:39 UTC por FLAVIO LOPES DE MORAIS

Atualizado: 2022-12-14 09:02:06 UTC por FLAVIO LOPES DE MORAIS