



Data Science Academy

www.datascienceacademy.com.br

Design e Implementação de Data
Warehouses

Enterprise Data Hub

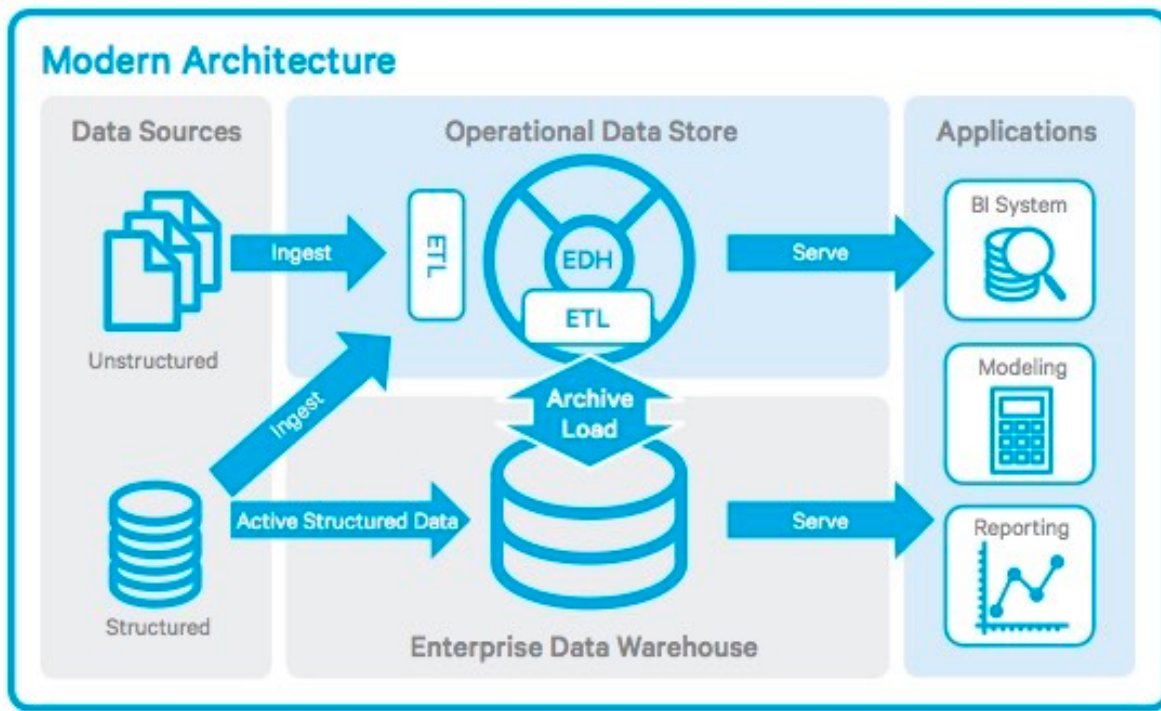
Um Enterprise Data Hub é um modelo de gerenciamento de Big Data que usa uma plataforma Hadoop como o repositório de dados central. O objetivo de um Enterprise Data Hub é permitir que a empresa tenha uma fonte de dados centralizada e unificada que possa fornecer rapidamente informações a diversos usuários de negócio, apoiando a tomada de decisão. Todos os aplicativos de analytics se conectam ao EDH para extrair as informações para análise.

O Big Data tornou-se a palavra de ordem nos últimos anos e muito mais ainda está por vir. Mas o grande desafio agora, é maximizar a utilização dos dados e alavancar benefícios reais, impulsionando oportunidades de negócio. Cada vez mais as empresas estão se movendo para uma cultura orientada a dados.

A introdução de um Enterprise Data Hub no cerne da arquitetura de informação de uma empresa, promove a centralização de todos os dados, em todos os formatos, disponíveis para todos os usuários de negócio, com total fidelidade e segurança e custo até 99% menor por Terabyte, em comparação com um Data Warehouse tradicional. O Hadoop foi desenvolvido para extrair o máximo de hardware de baixo custo.

Um Enterprise Data Hub serve como um repositório flexível para coletar e manter dados de forma ilimitada, seja para fins de conformidade ou para aplicações sofisticadas, como detecção de anomalias em tempo real. Ele acelera relatórios de analytics e aumenta a disponibilidade e acessibilidade dos dados para as atividades que suportam o crescimento do negócio e pode fornecer um quadro completo de operações para permitir a inovação de processos.

O conceito de EDH (Enterprise Data Hub), não chega a ser exatamente novo, pois já existem os EDWs (Enterprise Data Warehouses). O desafio está em mover a empresa de um modelo tradicional de gestão de dados, para um modelo voltado ao Big Data e suas particularidades como volume, velocidade e variedade. Os principais defensores do EDH são a Cloudera e a MapR. O gráfico abaixo mostra um exemplo de arquitetura moderna de dados, considerando o EDH e o EDW.



O Enterprise Data Hub inclui:

Reservatório de Dados ou “Data Lake”: Coleta de dados brutos que antes tinham alto custo para armazenamento e processamento. Os dados de diferentes fontes são gerenciados e governados no Data Lake, que também pode atuar como um arquivo online para dados acessados com menos frequência.

Refino de dados: Otimizar o processo de integração de diversos tipos de dados de várias fontes para descobrir as relações. Analisar, limpar, transformar e integrar dados.

Exploração do Big Data: Realizar análises investigativas em grandes volumes de dados de valor desconhecido, aplicando uma combinação de Machine Learning, estatísticas e técnicas de análise de gráfico SQL-on-Hadoop para descobrir novos conhecimentos e melhorar a análise operacional, tais como detecção de anomalia e sistemas de recomendação.

Fácil acesso aos dados: Os mais variados tipos de dados podem ser facilmente acessados em um EDH, garantindo uma fonte única para o trabalho de analytics.

Armazenamento de dados em formato nativo: Talvez uma das principais vantagens do EDH. Ao iniciar o trabalho de analytics, há a garantia que os dados estão em seu estado bruto, evitando distorção no processo de análise.

O Big Data está se consolidando e trazendo novos conceitos ou mudando conceitos existentes. À medida que as empresas adotarem uma cultura orientada a dados, veremos o Enterprise Data Hub no centro de qualquer estratégia de Big Data e Analytics. O DW pode ser um dos componentes desta arquitetura.

Fontes:

Cloudera Enterprise Data Hub

<https://www.cloudera.com/products/enterprise-data-hub.html>

Enterprise Data Hub

<https://mapr.com/solutions/enterprise/enterprise-data-hub/>

The Enterprise Data Hub and The Modern Information Architecture

https://www.microstrategy.com/Strategy/media/downloads/training-events/microstrategy-world/2014-vegas/MSTRWorld2014_T4_S2_The-Enterprise-Data-Hub-and-the-Modern-Information-Architecture.pdf?ext=.pdf

Enterprise Data Hub

<http://www.cienciaedados.com/enterprise-data-hub/>