

Ciência de Dados

Data-Centric Manifesto

Paradigma, Ideias Básicas:

Os Dados são o Centro do Universo.
Aplicativos são efêmeros.
São Visitantes Opcionais dos Dados.

A maior barreira na mudança de Paradigmas, é o Modelo Mental das pessoas envolvidas, e não tanto as Tecnologias.

Comparações de abordagens dos Dados:

APPLICATION CENTRIC

Alto custo de mudanças
Dados atados aos aplicativos
Cada projeto com seus dados
Dados diversificados fragmentados
Integração mais trabalhosa
Integração externa mais difícil
Maior risco de redundâncias
Dados complementares

DATA-CENTRIC

Custos mais razoáveis e estáveis
Dados são recursos independentes
Dados geralmente pré-existentes
Dados com integração planejada
Integração com custos mais baixos
Integração externa mais natural
Redundâncias são evitadas
Prioridade para dados corporativos

Data driven

É um adjetivo que qualifica processos orientados por dados, ou seja, embasados na coleta e análise de informações. No mundo dos negócios, significa colocar os dados no centro da tomada de decisão e do planejamento estratégico, buscando fontes confiáveis ao invés de gerir a empresa por intuição.

O termo data driven pode ser traduzido para “orientado por dados” e tem origem no conceito de ciência de dados.

Basicamente, essa ciência multidisciplinar usa métodos científicos, processos e algoritmos para extrair conhecimento de dados estruturados e não estruturados.

A ideia é usar a análise computacional (analytics) de grandes volumes de dados (Big Data) para solucionar problemas e obter insights, valendo-se de tecnologias em inteligência artificial e machine learning.

Administração de dados

Desenvolver e administrar de modo centralizado as estratégias, procedimentos e práticas para o processo de gerência dos recursos de dados e aplicativos, incluindo planos para sua definição, padronização, organização, proteção e utilização.

De acordo com a Data Management Association (DAMA), administração de dados é "o desenvolvimento e execução de arquiteturas, políticas, práticas e procedimentos que coordenam a estrutura de dados completa de uma ou mais organizações.