



PUC Minas
Virtual

Fundamentos de DataOps

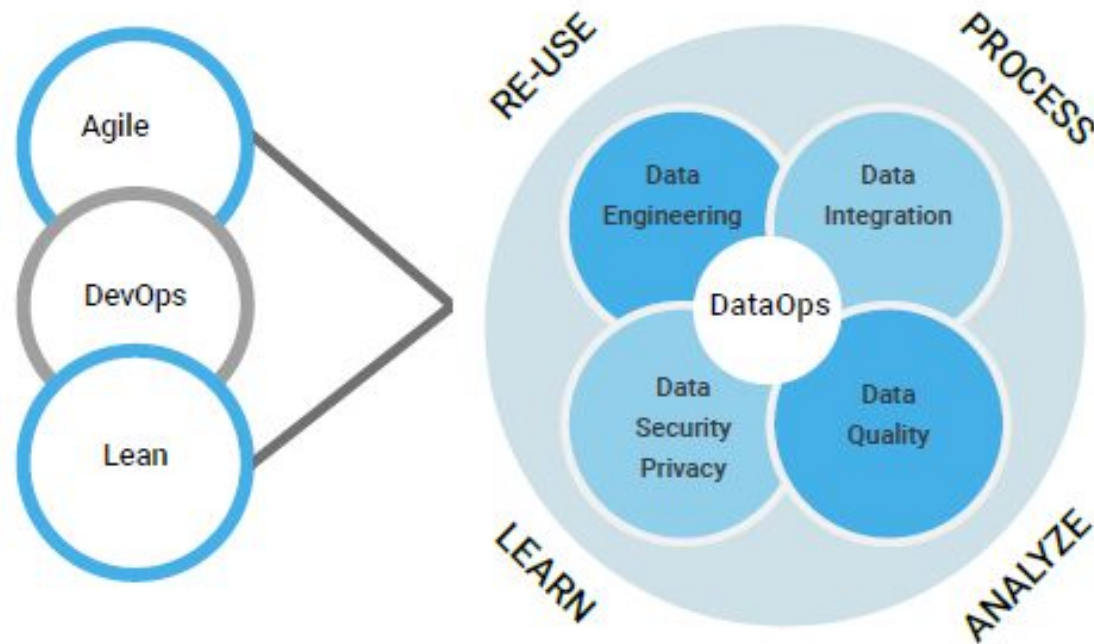
Renan Santos Mendes



PUC Minas
Virtual

Princípios de DataOps

Fundamentos de DataOps



Fonte: Intro to DataOps – Whitepaper – AWS

Fundamentos de DataOps

- Uma definição:

*“É um conjunto de **práticas, processos e tecnologias** que combina uma **perspectiva integrada e orientada a processos sobre dados** com **automação** e métodos de engenharia de software ágil para melhorar a **qualidade, velocidade e colaboração** além de promover uma cultura de **melhoria contínua** na área de análise de dados”*

Fundamentos de DataOps

Princípios:

- Satisfação contínua do cliente;
- Foco em análises relevantes;
- Abraçar a mudança;
- Interações diárias;
- Redução do heroísmo;
- Ter reprodutibilidade;
- Aplicar monitoramento;
- Melhorar o cycle time.

Fundamentos de DataOps

Semelhanças e diferenças com DevOps:

- Semelhanças:
 - Uso da metodologia ágil aplicada na gestão de projetos;
 - Entregas incrementais e com valor;
- Diferenças:
 - Objetivos diferentes: software X dados;
 - Fluxo de trabalho;
 - Testes de validação.

Fundamentos de DataOps

Motivos para ter uma cultura DataOps:

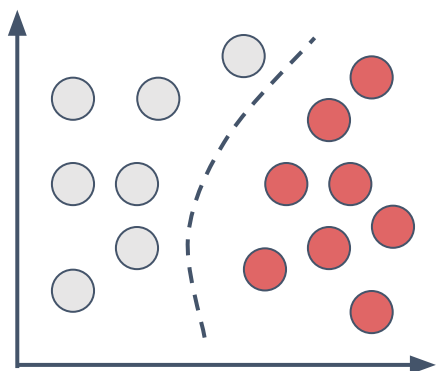
- Processos mais ágeis;
- Insights em tempo real;
- Democratização das informações;
- Foco estratégico;
- Alinhamento entre negócios e TI.

Aplicação: detectar *data drift* em um *dataset*

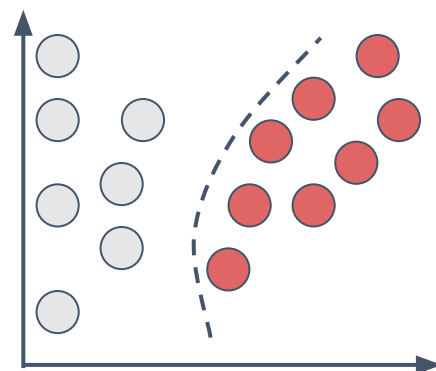
Fundamentos de DataOps

- Podemos aplicar uma validação no *dataset* usado para treinar;
- É analisado há uma **diferença significativa** em alguma coluna;
- Detecção de diferença = *data drift*;
- *Data drift* é um fenômeno em que os dados usados para treinar um modelo de ML **mudam com o tempo**;
- Isso pode levar a uma **diminuição nas métricas** do modelo.

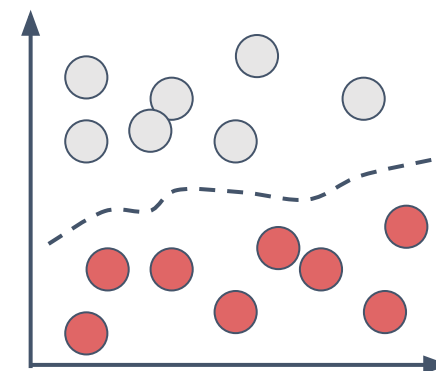
Fundamentos de DataOps



Dados
originais



*Data
Drift*



*Concept
Drift*

Fundamentos de DataOps

Diferenças entre *data drift* e *concept drift*:

- ***Data drift***: Mudança na distribuição dos dados;
- ***Concept drift***: Mudança na relação entre os dados e a variável alvo;
- **Resolução**: O *data drift* pode ser resolvido atualizando os dados do modelo. O *concept drift* pode ser resolvido refazendo o modelo.

Fundamentos de DataOps

Dicas para lidar com o *data drift* e o *concept drift*:

- **Monitore os dados:** monitorar os dados usados para treinar os modelos para detectar sinais de *drift*.
- **Atualize os dados regularmente:** Os modelos devem ser atualizados regularmente com novos dados;
- **Refaça os modelos:** Se o *concept drift* for severo, o modelo pode precisar ser refeito completamente.



PUC Minas
Virtual