

DEVOP E DATAOPS: VANTAGENS E DESVANTAGENS DAS MODERNAS TÉCNICAS DE ENGENHARIA DE SOFTWARE EM PROJETOS DE ANÁLISE E CIÊNCIA DE DADOS

Ciências Da Computação, Ciências Exatas e da Terra, Volume 27 - Edição 126 SET/23 SUMÁRIO / 14/09/2023

DEVOPS AND DATAOPS: ADVANTAGES AND DISADVANTAGES OF MODERN SOFTWARE ENGINEERING TECHNIQUES IN DATA ANALYSIS AND DATA SCIENCE PROJECTS

REGISTRO DOI: 10.5281/zenodo.8343203

Jessica Alves Rodrigues Macedo^[1]

Thales Kroth de Souza^[1]

Wesley de Sá Teles^[2]

Daniel Clarismundo Borges^[3]

Resumo

DevOps é um acrônimo para designar o conjunto de práticas de Engenharia de *Software* que combina desenvolvimento de *software* (Dev) e operações de Tecnologia da Informação (Ops). Seu objetivo é encurtar o ciclo de vida de desenvolvimento de sistemas fornecendo entrega contínua com alta qualidade de *software* combinando com o uso de metodologias ágeis. A aplicação dos conceitos e técnicas do DevOps na análise e ciência de dados é conhecido como DataOps, um acrônimo para uma abordagem de gerenciamento de dados (Data) com o objetivo melhorar a eficiência, agilidade e qualidade de um ciclo de vida de dados (Ops). Este trabalho apresenta as vantagens das modernas técnicas de engenharia de *software* por meio do DevOps e DataOps e qual abordagem utilizar segundo o contexto.

Palavras-chave: DevOps. DataOps. Engenharia de *Software*. Análise de Dados. Ciência de Dados.

Abstract

DevOps is an acronym for the set of Software Engineering practices that combine software development (Dev) and Information Technology operations (Ops). The principle is to shorten the systems development life cycle by providing continuous delivery with high-quality software and combining development with agile methodologies. The application of DevOps concepts and techniques to data science and analysis is known as DataOps, an acronym for managing data (Data) to improve the efficiency, agility, and quality of a data lifecycle (Ops). This paper presents the advantages of modern software engineering approaches through DevOps and DataOps and which approach to use according to the context.

Keywords: DevOps. DataOps. Software Engineering. Analytics. Data Science.

1 INTRODUÇÃO

A Engenharia de *Software* é uma área da Ciência da Computação que se dedica a projetar, desenvolver, testar e manter *softwares* com qualidade, tendo ênfase na aplicação de técnicas e metodologias que fortalecem a efetividade, eficiência e confiabilidade dos produtos desenvolvidos. (SOMMERVILLE, 2011)

Trata-se de um processo iterativo e incremental que envolve a utilização de uma série de atividades, tais como análise de requisitos, projeto, implementação, teste, integração e manutenção de *software*, com o objetivo de criar soluções de *software* eficazes e eficientes, que atendam às necessidades dos usuários. (PRESSMAN e MAXIM, 2015)

Além disso, a Engenharia de *Software* tem como objetivo promover a reutilização de *software*, a padronização de processos e a melhoria contínua da qualidade dos produtos de *software*. (LEACH, 2018)

O conjunto de práticas da Engenharia de *Software* que visam aproximar e integrar as equipes de desenvolvimento e operações para melhorar a velocidade, qualidade e eficiência do processo de entrega de *software* é denominado DevOps (BASS, 2015)

Consiste numa abordagem que une os processos, as ferramentas e as culturas de Desenvolvimento (Dev) e Operações (Ops) em um ciclo contínuo, desde o planejamento e o desenvolvimento até o monitoramento e o suporte pós-implementação.

É uma prática que envolve a colaboração entre as equipes de desenvolvimento de *software* e as de operações, visando à entrega contínua e automatizada de *software* de alta qualidade (DAVIS e DANIELS, 2016).

Como resultado, a DevOps permite que as equipes de TI entreguem soluções rapidamente, com maior qualidade e segurança. Neste contexto o DevOps pode ser aplicado para projetos de desenvolvimento de *software* a fim de acelerar a entrega de soluções mais confiáveis e garantindo que estas soluções sejam escaláveis e resilientes.

Isso é possível, porque o DevOps se concentra em automatizar o processo de desenvolvimento e implantação de *software*, fornecendo aos desenvolvedores as ferramentas necessárias para produzir soluções de maneira mais rápida e eficiente sem despendar muito tempo com processos manuais. (HUMBLE e KIM, 2018)

Por outro lado, a análise e ciência de dados é o processo de extrair informações úteis a partir de grandes conjuntos de dados.

É uma combinação de habilidades em estatística, programação, visualização e comunicação, que permite aos profissionais explorar e compreender dados, identificar padrões e tendências, bem como prever comportamentos futuros.

Essa disciplina envolve “transformar dados em ação”, usando métodos científicos para obter *insights* e orientar a tomada de decisões em negócios e outras áreas. (PROVOST E FAWCETT, 2013; BLUM, HOPCROFT, e KANNAN, 2020)

Os projetos de análise e ciência de dados apresentam diversos desafios, desde a coleta e armazenamento de dados até a sua análise e interpretação. A falta de dados precisos e confiáveis, a integração de diferentes fontes de dados e a falta de infraestrutura adequada para gerenciamento de dados, são alguns dos desafios enfrentados no início do projeto. (KIMBALL & CARSERTA, 2004)

Na fase de análise, o desafio é garantir a qualidade dos dados, definir as técnicas e algoritmos de análise adequados e obter resultados precisos e úteis. Por fim, a interpretação dos resultados e sua transformação em conhecimentos úteis para a tomada de decisão é outro grande desafio. Além disso, é preciso considerar aspectos éticos e de privacidade no tratamento dos dados. (CADY, 2017; PENG, CAFFO e LEEK, 2018)

Dado que a abordagem DevOps visa aprimorar a colaboração e comunicação entre as equipes de desenvolvimento de *software* e operações, buscando a automação, integração e entrega contínua de *softwares*, sua utilização pode ajudar a enfrentar os desafios nos projetos de análise e ciência de dados ao promover uma cultura colaborativa e de compartilhamento de conhecimento, reduzindo a lacuna entre desenvolvedores e cientistas de dados.

A abordagem DevOps também pode ser usada para ajudar a garantir que as soluções de análise de dados sejam seguras, confiáveis e de alta qualidade ao incluir processos de verificação e validação contínuos, em conjunto com as melhores práticas de segurança e gerenciamento de configurações.

Entretanto a aplicação dos conceitos e técnicas do DevOps em projetos de análise e ciência de dados é conhecido como DataOps, um acrônimo para uma abordagem para gerenciamento de dados (Data) com o objetivo de melhorar a eficiência, agilidade e qualidade de um ciclo de vida de dados (Ops).

O DataOps combina metodologias ágeis com automação, infraestrutura e integração contínua, para melhorar a velocidade e qualidade na criação, manutenção e uso de dados (BASS, 2015).

Seu objetivo é permitir que as equipes de tecnologia criem e implementem soluções de dados de forma mais rápida e com menos erros visando a qualidade, atendendo às necessidades de negócios com mais eficiência.

O DataOps também enfatiza a colaboração entre equipes, incluindo desenvolvedores, engenheiros de dados, administradores de sistemas e analistas de negócios, para que todos trabalhem juntos em projetos de dados a alcançar resultados melhores e mais rápidos. (TREWING, 2021).

DevOps e DataOps são conceitos relacionados, mas com focos diferentes. Enquanto DevOps tem como objetivo principal a integração entre as equipes de desenvolvimento e operações de TI, visando a entrega contínua de *software* de alta qualidade, DataOps é um conceito mais recente que busca aplicar os mesmos princípios do DevOps à área de dados. (BASS, WEBER e ZHU, 2019)

O objetivo do DataOps é acelerar a entrega de soluções de dados de alta qualidade, promovendo uma cultura de colaboração, automação e monitoramento contínuo em todas as etapas do ciclo de vida dos dados. Em resumo, enquanto DevOps se concentra em *software*, DataOps se concentra em dados. (BERGH, 2019).

Este trabalho visa apresentar as vantagens que as abordagens DevOps e DataOps podem trazer para a análise e ciência de dados a partir da revisão de obras nas áreas de Engenharia de *Software*, DevOps, DataOps, Análise e Ciência de Dados.

Está organizado da seguinte maneira: a seção 2 apresenta os principais tópicos encontrados na literatura interdisciplinar; a seção 3 detalha o método de pesquisa adotado; a seção 4 apresenta uma discussão sobre a análise realizada; a seção 5 apresenta as considerações finais do trabalho, apresentando recomendações para o uso do DevOps e DataOps em projetos de análise e ciência de dados.

1.1. METODOLOGIA DA PESQUISA

A abordagem de pesquisa adotada neste artigo é a qualitativa, sendo ideal quando queremos descrever nosso objeto de estudo com mais profundidade e quando a influência do pesquisador é considerada fundamental

(MASCARENHAS, 2018).

Trata-se de uma pesquisa bibliográfica acerca do assunto com revisão sistemática, segundo o procedimento técnico realizado que buscou pesquisar, selecionar, avaliar, sintetizar e relatar os achados sobre uma determinada pergunta ou tópico (ROEVER, 2020).

Com base na pergunta norteadora: “Quais vantagens as modernas técnicas de engenharia de *software*, por meio de DevOps e DataOps, trazem para lidar com os desafios em projetos de análise e ciência de dados?”, buscou-se a seleção de livros referência nas áreas de estudos em questão.

A partir de um levantamento preliminar, realizado entre janeiro e julho de 2023, foram selecionados 50 livros com base na experiência de trabalho de campo do autores e através da aplicação de descritores com base nas áreas de estudo para a seleção dos livros, a saber: Engenharia de *Software*, DevOps, DataOps, Análise e Ciência de Dados.

Após análise do conteúdo de cada um dos livros selecionados (KRIPPENDORFF, 2004) buscou-se a ordenação dos por meio de agrupamento em função de seus significados e semelhanças (CORBIN e STRAUSS, 2015) nas áreas de estudo.

“O quão específico este livro aborda uma determinada tecnologia ou linguagem de programação em relação aos livros outros selecionados?”, foi o primeiro critério utilizado para remoção de 26 livros numa primeira etapa. A especificidade de uma tecnologia ou linguagem de programação reduziria sua a colaboração para responder à pergunta norteadora deste artigo.

Foram selecionados 24 livros, agrupados segundo as áreas de estudos: 08 na área de DevOps, 08 na área de Análise e Ciência de Dados, 04 na área de Engenharia de *Software* e 04 na área de DataOps.

Os 24 livros selecionados foram brevemente resumidos no referencial teórico deste artigo e os achados discutidos à luz da pergunta norteadora.

1.2. OBJETIVOS

O principal objetivo deste trabalho é explorar e apresentar as vantagens que as abordagens DevOps e DataOps podem oferecer para projetos de análise e ciência de dados. Isso inclui a melhoria da colaboração entre equipes, a aceleração da entrega de soluções de dados de alta qualidade e a promoção de uma cultura de automação e monitoramento contínuo. Além disso, o trabalho busca fornecer uma revisão da literatura nas áreas de Engenharia de Software, DevOps, DataOps, Análise e Ciência de Dados para embasar a discussão.

1.3. JUSTIFICATIVA

A análise e ciência de dados têm se tornado fundamentais para as organizações em um mundo cada vez mais orientado por dados. No entanto, esses projetos enfrentam diversos desafios, desde a coleta e qualidade dos dados até a análise e interpretação dos resultados.

Nesse contexto, as abordagens DevOps e DataOps surgem como soluções promissoras para otimizar e acelerar o ciclo de vida de dados, promovendo a colaboração entre equipes e a entrega de soluções de alta qualidade. Compreender as vantagens dessas abordagens é essencial para profissionais e organizações que desejam adotar práticas mais eficazes na área de análise e ciência de dados.

Portanto, este trabalho busca fornecer uma base sólida de conhecimento e recomendações para o uso do DevOps e DataOps nesse contexto, contribuindo para o avanço da área e melhorias nos resultados dos projetos de análise e ciência de dados.

2 FUNDAMENTAÇÃO TEÓRICA

O referencial teórico para este artigo é composto pela revisão de livros selecionados que foram ordenados e agrupados segundo suas áreas de interesse: Engenharia de *Software*, DevOps, DataOps, Análise e Ciência de Dados.

Buscou-se com esta seleção, livros que permitissem compreender os desafios dos projetos de desenvolvimento de *software* e de análise e ciência de dados, assim como as vantagens que a utilização das abordagens DevOps e DataOps trazem para enfrentar estes desafios.

2.1. ENGENHARIA DE SOFTWARE

A área de Engenharia de *Software* apresenta vários desafios no projeto de desenvolvimento de *software* (SOMMERVILLE, 2011; PRESSMAN e MAXIM, 2015; LEACH, 2018; TSUI, KARAM e BERNAL, 2022) e segundo os autores, alguns dos principais são:

2.1.1 Envolvimento do usuário: envolvê-los durante todo o ciclo de vida do *software* é fundamental para atender suas necessidades e expectativas. No entanto, isso pode ser difícil devido a problemas como a disponibilidade do usuário e as habilidades necessárias para a participação;

2.1.2 Mudanças constantes: alterações no ambiente de negócios, necessidades do usuário e avanços tecnológicos podem levar a mudanças no escopo e nos requisitos do projeto. Essas modificações podem dificultar a manutenção da integridade do projeto e a entrega dentro do prazo;

2.1.3 Gerenciamento de requisitos: é importante que os requisitos sejam gerenciados de forma eficaz para garantir que as necessidades do usuário sejam atendidas e o escopo do projeto seja mantido. Fatores como o tamanho do projeto, a complexidade do *software* e as necessidades em constante mudança dos usuários podem impactar o gerenciamento de requisitos;

2.1.4 Qualidade do *software*: garantir a qualidade do *software* é fundamental para garantir a satisfação do usuário e a eficácia do produto. Contudo, manter esta qualidade pode ser um desafio em ambientes de mudanças constantes;

2.1.5 Produtividade: é uma questão-chave em qualquer projeto de engenharia de *software*, pois pode afetar o custo e o prazo do projeto. Melhorar a produtividade pode ser difícil devido a fatores como a complexidade do projeto e o tamanho da equipe.

2.1.6 Comunicação: é fundamental que seja eficaz para garantir que todos os membros da equipe tenham uma compreensão clara do projeto e de seus requisitos. No entanto, a comunicação pode ser afetada por diversos fatores como: distância física, diferenças culturais ou linguísticas entre os membros da equipe;

2.1.7 Gerenciamento de projetos: é importante para garantir que o projeto seja entregue dentro do prazo e do orçamento. Fatores como a complexidade do *software*, envolvimento de muitos *stakeholders* e a incerteza associada ao desenvolvimento de *software* podem impactar o gerenciamento do projeto;

2.1.8 Desenvolvimento ágil: a adoção de metodologias ágeis pode ajudar a acelerar o desenvolvimento de *software* e a melhorar a qualidade do produto. A adaptação ao novo modelo de trabalho pode ser desafiadora para gestão de equipes;

2.1.9 Segurança: é fundamental para proteger os dados dos usuários e as informações sensíveis do negócio, sendo sempre um item de grande importância e complexidade.

2.2. ANÁLISE E CIÊNCIA DE DADOS

De forma análoga, os desafios que podem surgir durante um projeto de Análise e Ciência de Dados são específicos do ciclo de vida dos dados (KIMBALL & CARSERTA, 2004; PENG e MATSUI, 2016; SKIENA, 2017; KELLEHER e TIERNEY, 2018) e, segundo os autores, alguns dos principais são:

2.2.1 Aquisição de dados: obter dados de fontes confiáveis e relevantes é um dos maiores desafios de um projeto de ciência de dados. É importante garantir que os dados sejam precisos, completos e de alta qualidade antes de serem utilizados para análise;

2.2.2 Limpeza e preparação de dados: pode ser uma tarefa trabalhosa e demorada. Alguns conjuntos de dados incluem dados ausentes, inconsistentes ou irrelevantes que precisam ser corrigidos ou removidos antes de serem utilizados para análise;

2.2.3 Modelagem de dados: escolher o modelo de dados correto é importante para garantir que as informações sejam apresentadas de maneira clara e significativa. É importante considerar as limitações e as hipóteses do modelo antes de usá-lo para tomar decisões;

2.2.4 Visualização de dados: é importante para ajudar a interpretar os resultados da análise de dados. É necessário escolher a melhor forma de visualização para representar de maneira clara, objetiva e de fácil entendimento;

2.2.5 Modelagem preditiva: envolve o uso de algoritmos para prever o resultado futuro com base em dados passados. Escolher o algoritmo correto é importante para garantir resultados precisos e relevantes;

2.2.6 Implementação e manutenção: são itens dentro de um sistema de ciência de dados que podem ser desafiadores. É importante garantir que o sistema esteja sempre atualizado, disponível e funcionando corretamente.

2.2.7 Tomada de decisão: a ciência de dados é uma ferramenta para ajudar a tomar decisões, mas as decisões finais ainda precisam ser realizadas por pessoas, desta forma, faz-se necessário garantir que as decisões sejam baseadas em dados confiáveis e relevantes;

2.2.8 Integração de dados: integrar dados de diferentes fontes e sistemas pode ser desafiador e requer habilidades em gerenciamento de dados. Além disso, é importante garantir a qualidade dos dados integrados, já que os resultados da análise de dados dependem da qualidade dos dados;

2.2.9 Escolha de algoritmos: A escolha de algoritmos apropriados para a análise de dados pode ser difícil, especialmente se houver uma grande quantidade de algoritmos disponíveis. Contudo, é importante ter conhecimento sobre como os algoritmos funcionam e suas limitações para que seja realizado uma escolha assertiva para o problema em questão;

2.2.10 Validação de modelos: para modelos de análise de dados é importante garantir que eles sejam precisos e confiáveis. Isso pode incluir a avaliação da acurácia do modelo em diferentes cenários.

2.2.11 Comunicação de resultados: informar os resultados da análise de dados de maneira clara e concisa é fundamental para que os dados sejam compreendidos e utilizados de maneira eficaz. Isso pode incluir a apresentação de resultados em relatórios, painéis de dados interativos ou apresentações em reuniões.

2.2.12 Implementação: construir soluções baseadas em dados pode ser desafiador, especialmente se houver necessidade de integração com sistemas existentes ou se precisar escalar soluções para suportar grandes

volumes de dados. Além disso, é importante monitorar continuamente as soluções para garantir que elas estejam funcionando corretamente e sejam ajustadas à medida que a demanda mude.

2.3. DEVOPS E DATAOPS: UMA ANÁLISE CONTRASTIVA

DevOps é uma metodologia que visa a integração entre equipes de desenvolvimento e operações, buscando otimizar a colaboração, a entrega de valor, a qualidade e a segurança do *software* (HUMBLE & FARLEY, 2010; WALLS 2013).

Diante dos desafios enfrentados em projetos de Engenharia de *Software*, Análise e Ciência de Dados, o uso de práticas e ferramentas de DevOps pode contribuir para superar obstáculos como a complexidade, a falta de colaboração, a falta de automação, a falta de padronização e a falta de visibilidade do processo (KIM, HUMBLE, DEBOIS, WILLIS e FORSGREN, 2010).

Através de processos automatizados e padronizados, que promovem a colaboração e a comunicação constante entre as equipes, a abordagem DevOps pode ajudar a mitigar riscos e aumentar a eficiência do desenvolvimento e entrega de *software*. Além disso, a adoção de práticas de integração contínua, entrega contínua e monitoramento constante podem permitir uma maior visibilidade e *feedback* em tempo real, possibilitando uma tomada de decisão mais rápida e assertiva. (KIM, BEHR, SPAFFORD, 2013).

DevOps e DataOps são duas abordagens diferentes, mas que possuem muitas semelhanças. Enquanto o DevOps é focado na integração contínua e na entrega contínua de *software*, o DataOps é voltado para o gerenciamento de dados, desde a coleta até a análise (BERGH, 2019).

Enquanto DevOps visa automatizar e otimizar os processos de desenvolvimento, teste e implantação, o DataOps procura automatizar a captura, a limpeza e a preparação de dados para análise, bem como monitorar e gerenciar todo o fluxo de dados. Ambas as disciplinas compartilham uma filosofia de colaboração e automação, com o objetivo de fornecer soluções de alta qualidade de forma mais rápida e eficiente (THUSOO e SARMA, 2019).

A abordagem DataOps tem sido proposta como uma evolução da abordagem DevOps, especialmente para lidar com projetos de Análise e Ciência de Dados. A principal justificativa para essa evolução é que projetos de análise de dados geralmente lidam com conjuntos de dados maiores e mais complexos e têm requisitos mais específicos em termos de qualidade, confiabilidade e segurança (THUSOO e SARMA, 2019).

A colaboração estreita entre os cientistas de dados, desenvolvedores e operadores que a abordagem DevOps promove pode não ser suficiente para lidar com a complexidade de projetos de análise de dados (WALLS, 2013).

DataOps tem como objetivo integrar a gestão de dados, análise e entrega em um ciclo de vida contínuo e automatizado, permitindo que as equipes trabalhem juntas de forma mais eficiente e segura. Ela enfatiza a colaboração e a comunicação entre as equipes, a automação de processos e a padronização de ferramentas e metodologias (BERGH, 2019).

Ao adotar a abordagem DataOps, as equipes podem se beneficiar de práticas de gerenciamento de dados mais avançadas, como a implementação de uma infraestrutura de dados segura e escalável, a definição de padrões para a coleta, limpeza e armazenamento de dados e o monitoramento contínuo da qualidade e desempenho dos dados (THUSOO e SARMA, 2019).

3 RESULTADOS E DISCUSSÕES

DataOps e DevOps representam abordagens modernas na Engenharia de *Software*, embora com focos distintos. Essa diferenciação fica evidente ao analisarmos as vantagens que cada uma dessas abordagens oferecem ao

enfrentar os desafios propostos e ao compará-las de acordo com alguns critérios.

3.1. PRINCIPAIS VANTAGENS DA ABORDAGEM DEVOPS

Segundo a literatura (LOUKIDES, 2012; BASS, 2015; DAVIS e DANIELS, 2016), as principais vantagens da abordagem DevOps para projetos de desenvolvimento de *software* são:

3.1.1. Maior frequência de implementações e entregas de *software*, o que permite que as empresas acompanhem a demanda do mercado e inovem com mais frequência;

3.1.2. Redução do tempo de desenvolvimento e entrega de software, permitindo a resolução mais rápida de problemas e reduzindo os custos operacionais;

3.1.3. Maior colaboração entre equipes de desenvolvimento e operações, permitindo que os desenvolvedores recebam feedbacks rapidamente com objetivo de melhorar o processo;

3.1.4. Melhor qualidade de software ao permitir testes automatizados e implementações de correções de forma ágil;

3.1.5. Redução do risco de falhas ao estimular que as equipes identifiquem antecipadamente problemas que só seriam descobertos em produção;

3.1.6. Redução do tempo de inatividade ao estimular que os desenvolvedores corrijam problemas e minimizem os impactos negativos nas operações da empresa;

3.1.7. Maior eficiência operacional fomentado a automação de muitos processos e permitindo que as equipes se concentrem em atividades de maior valor;

3.1.8. Melhor segurança estimulando a implementação de práticas de segurança em todas as etapas do processo de desenvolvimento e entrega;

3.1.9. Melhor controle de versionamento ao permitir que as equipes controlem as versões do software para evitar conflitos entre diferentes versões;

3.1.10. Maior alinhamento entre as equipes de desenvolvimento e operações e a estratégia da empresa, permitindo que todos estejam trabalhando em direção aos mesmos objetivos.

3.2. PRINCIPAIS VANTAGENS DA ABORDAGEM DATAOPS

Já na abordagem DataOps (BERGH, 2019; ATWAL, 2019; THUSOO e SARMA, 2019; TREWIN, 2021), as principais vantagens para projetos de análise e ciência de dados são:

3.2.1. Maior colaboração entre as equipes de dados, desenvolvimento e operações. Isso ajuda a evitar silos de informação, garantindo que todas as partes interessadas estejam alinhadas e trabalhando juntas em prol do sucesso do projeto;

3.2.2. Aceleração da entrega de dados ao permitir que as equipes de dados entreguem dados de forma mais rápida e eficiente, reduzindo o tempo necessário para o tratamento de dados;

3.2.3. Qualidade dos dados melhorada ao promovê-la ao longo de todo o ciclo de vida do projeto, ajudando a garantir que os dados sejam precisos, confiáveis e relevantes;

3.2.4. Aumento da eficiência das equipes de dados, permitindo que elas trabalhem com mais rapidez e eficiência, reduzindo os gargalos do processo;

3.2.5. Maior agilidade ao permitir que as equipes de dados respondam às mudanças e demandas do negócio;

3.2.6. Redução de custos ao garantir que as equipes de dados não percam tempo com tarefas manuais ou retrabalho;

3.2.7. Maior transparência ao fornecer visibilidade para todo o processo de gerenciamento de dados, desde a ingestão até a entrega;

3.2.8. Flexibilidade para que as equipes se adaptem rapidamente a novas demandas e requisitos;

3.2.9. Facilidade de manutenção simplificando o processo de manutenção de dados, permitindo que as equipes de dados mantenham as informações atualizadas e relevantes;

3.2.10. Melhores abordagens garantindo que as equipes de dados obtenham insights precisos e relevantes e permitindo que as empresas tomem decisões mais informadas e eficazes.

3.3. DIFERENÇAS IMPORTANTES ENTRE AS ABORDAGENS

Embora DataOps compartilhe muitos princípios e práticas com DevOps, há algumas diferenças importantes entre as abordagens. Comparando as vantagens com base em critérios é possível destacar essas diferenças. A seguir são apresentados alguns desses critérios de comparação:

3.3.1. Objetivo: enquanto o DevOps foca em garantir a entrega rápida e confiável de software, o DataOps foca em garantir a entrega rápida e confiável de dados;

3.3.2. Equipes: enquanto o DevOps envolve principalmente equipes de desenvolvimento e operações, o DataOps envolve equipes de análise de dados, cientistas de dados e engenheiros de dados;

3.3.3. Dados como produto: no DataOps, os dados são tratados como um produto e, portanto, têm seus próprios ciclos de vida, qualidade e governança.

3.3.4. Ferramentas: embora muitas ferramentas de DevOps possam ser usadas em DataOps, existem ferramentas específicas para DataOps, como Apache Airflow, Databricks e Dataform. Essas ferramentas foram desenvolvidas especificamente para lidar com os desafios relacionados à manipulação, preparação, governança e execução de tarefas de processamento de dados no contexto de DataOps.

3.3.5. Fluxo: Em DataOps, o fluxo de trabalho difere do fluxo de trabalho em DevOps, especialmente em relação às fases de desenvolvimento e testes. Em DataOps, essas fases são mais focadas em preparar e limpar os dados, pois a qualidade e disponibilidade dos dados são aspectos críticos para garantir resultados precisos e confiáveis nas análises e aplicações de dados, em contrapartida em DevOps, o foco é mais voltado para a integração contínua, entrega contínua e automação de processos de desenvolvimento e operações de software. A abordagem de DataOps valoriza o tratamento adequado dos dados desde o início do processo, o que é fundamental para assegurar a qualidade dos resultados finais.

3.3.6. Testes: o teste em DataOps é mais complicado do que em DevOps, uma vez que é preciso garantir que os dados sejam precisos e estejam sendo usados corretamente, impactando significativamente a automação;

3.3.7. Processamento distribuído: em DataOps, o processamento distribuído de dados é mais comum, pois muitos projetos envolvem grandes quantidades de dados.

3.3.8. Governança de dados: a governança de dados é muito importante na abordagem DataOps, uma vez que os dados precisam ser de alta qualidade, precisão e conformidade.

Quadro 01 – Resumo comparativo entre principais diferença

Característica	DevOps	DataOps
Foco	Desenvolvimento e operação de <i>software</i>	Coleta, análise e gerenciamento de dados
Objetivos	Automatizar e integrar os processos de desenvolvimento e operação de <i>software</i> .	Melhorar a eficiência e eficácia dos processos de coleta, análise e gerenciamento de dados.
Ferramentas	Ferramentas de automação, gerenciamento de configuração e controle de versão.	Ferramentas de análise e gerenciamento de dados e colaboração.
Equipe envolvida	Desenvolvimento de <i>software</i> , operações de TI e qualidade.	Desenvolvimento de <i>software</i> , operações de TI, análise de dados e negócios.

Fonte: Autores, 2023.

Na abordagem de especificidade sobre a atuação das equipes que participam de projetos onde constam DevOps e DataOps, enquanto a primeira transforma o desenvolvimento de *softwares* com novas atualizações de *software*, a segunda transforma a área de *data science* e *analytics*, essencialmente para com a necessidade dos negócios modernos mais emergentes e líquidos (LIMA, 2023).

Se DevOps possui aspecto para a qualidade, versões mais otimizadas e um trabalho contínuo para com o produto que impacta o cliente final, DataOps parte da corrente sobre o tratamento dos dados e o panorama sobre sua indústria, fazendo com que a tecnologia tenha uma aplicação recorrente de escolha no ciclo coleta, análise e gerenciamento, oportunizando detalhes afins de operações que envolvam tecnologia.

O foco em dados como um produto, processamento distribuído, fluxo de trabalho e governança de dados torna o DataOps essencial para projetos de análise e ciência de dados. Por outro lado, o DevOps continua sendo a melhor opção para o desenvolvimento de *software*

4 CONSIDERAÇÕES FINAIS

Engenharia de *Software* é uma área com histórico de diversas abordagens para lidar com os desafios dos projetos de desenvolvimento de *software* em aplicações aos negócios, problemas estruturais tecnológicos e de necessidades para uma área específica.

Projetos de análise e ciência de dados compartilham desafios semelhantes com certas particularidades inerentes ao ciclo de vida de dados ao cerne da tecnologia tratada ou não.

A adoção de práticas de DevOps tem se mostrado essencial para superar esses desafios em projetos de engenharia de *software* com melhorias significativas na qualidade do *software*, redução de erros e maior eficiência operacional e de arquitetura tecnológica para o fluxo transacional de informações.

No entanto, para projetos de análise e ciência de dados, a abordagem DataOps é mais adequada, pois considera as particularidades do trabalho com dados, desde a coleta, preparação, processamento, tratamento, modelagem

até a entrega de conhecimento relevante na cadeia informacional.

O uso de tecnologias modernas e a integração entre as equipes de dados, desenvolvimento e operações são fundamentais para o sucesso desses projetos. DevOps e DataOps representam abordagens complementares que ajudam a enfrentar os desafios de projetos afins que mesclam-se em análise e na ciência de dados.

A adoção de abordagens DevOps e DataOps pode aumentar a eficiência, reduzir os custos e melhorar a qualidade de projetos de Engenharia de *Software*, análise, desenvolvimento, tecnologia e ciência de dados conjuntamente que colaboram com as organizações a se manterem competitivas e atrativas em um mercado cada vez mais exigente.

4.1. RECOMENDAÇÕES PARA FUTUROS ESTUDOS

Entre as orientações para estudos futuros em DevOps e DataOps, destacam-se as seguintes recomendações:

4.1.1. Aprofundar a compreensão dos princípios e práticas das abordagens em DevOps e DataOps, bem como suas aplicações em diferentes contextos e ambientes de desenvolvimento de *software* e análise de dados em linha de negócios para facilitar ora a ciência, ora a análise de informações e dados;

4.1.2. Analisar as ferramentas e tecnologias associadas à implantação contínua, integração contínua, monitoramento e gerenciamento de infraestrutura e ambientes de desenvolvimento e produção de dados;

4.1.3. Analisar as principais ferramentas de análise e visualização de dados, bem como as técnicas de processamento e preparação de dados para análise como limpeza, transformação, normalização e enriquecimento de dados;

4.1.4. Comparar dois fluxos distintos de transmissão de dados em um cenário para aperfeiçoar DevOps em operações de *software*;

4.1.5. Como o aperfeiçoamento dos processos de coleta, análise e gerenciamento de dados que envolvam DataOps caracterizam-se com as novas tecnologias de inteligência criativa ou artificial;

4.1.6. Se ferramentas tecnológicas que gerenciam dados podem ser replicadas em outros estudos envolvendo DataOps.

Essas recomendações podem auxiliar profissionais e pesquisadores a aprimorarem suas habilidades e conhecimentos em DevOps e DataOps, proporcionando oportunidades para aprofundamento em áreas específicas.

Além disso, considerando o avanço constante da tecnologia, esses estudos têm grande relevância para campos do conhecimento que demandam pesquisas mais aprofundadas. Eles também contribuem para o enriquecimento das informações e estudos nas áreas de engenharia de software e ciência de dados, ambas fundamentais no campo da tecnologia.

Com o aumento da complexidade das soluções de software e o crescente volume de dados, a capacidade de aplicar princípios do DevOps e DataOps torna-se essencial para o desenvolvimento e análise de sistemas eficazes e eficientes, abrindo portas para inovações e avanços significativos em projetos tecnológicos.

REFERÊNCIAS

ATWAL, Harvinder. **Practical DataOps: Delivering agile data science at scale**. Apress, 2019.

BASS, Len; WEBER, Ingo; ZHU, Liming. **DevOps: A software architect's perspective**. Addison-Wesley Professional, 2015.

BERGH, Christopher; BENGHIAT Gil; STROD Eran – **The DataOps Cookbook: Methodologies and Tools That Reduce Analytics Cycle Time While Improving Quality**. DataKitchen, 2019

BLUM, Avrim; HOPCROFT, John; KANNAN, Ravindran. **Foundations of data science**, Cambridge University Press, 2020.

CADY, Field, **The data science handbook**, John Wiley & Sons Inc, 2017.

CORBIN, Juliet M.; STRAUSS, Anselm. **Grounded theory research: Procedures, canons, and evaluative criteria. Qualitative sociology**, v. 13, n. 1, p. 3-21, 1990.

DAVIS, Jennifer; DANIELS, Ryn. **Effective DevOps: building a culture of collaboration, affinity, and tooling at scale**. “O’Reilly Media, Inc.”, 2016.

HUMBLE, Jez; KIM, Gene. **Accelerate: The science of lean software and devops: Building and scaling high performing technology organizations**. Portland: IT Revolution, 2018.

HUMBLE, Jez; FARLEY, David. **Continuous Delivery: Reliable Software Releases Through Build, Test, and Deployment Automation**. Addison-Wesley Professional, 2010.

KELLEHER, John D.; TIERNEY, Brendan. **Data science**, Cambridge: MIT Press, 2018.

KIM, Gene; BEHR, K.; SPAFFORD, G. **The Phoenix Project: A Novel About IT, DevOps, and Helping Your Business Win**. Portland: IT Revolution Press, 2013.

KIM, G., HUMBLE, J., DEBOIS, P., WILLIS, J., & FORSGREN, N., **The DevOps handbook: How to create world-class agility, reliability, & security in technology organizations**, Portland: IT Revolution Press, 2021.

KIMBALL, Ralph; CASERTA, Joe. **The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data**. 1. ed. Wiley, 2004.

KRIPPENDORFF, Klaus. **Content analysis: An introduction to its methodology**. Sage publications, 2018.

LEACH, Ronald J. **Introduction to software engineering**. CRC Press, 2018.

LOUKIDES, Mike. **What is DevOps?**. “O’Reilly Media, Inc.”, 2012.

MASCARENHAS, Sidnei A; **Metodologia Científica**, 2ª ed, São Paulo: Pearson Education do Brasil, 2018.

PENG, Roger D.; MATSUI, Elizabeth; **The Art of Data Science: A Guide for Anyone Who Works with Data**, Leanpub, 2016.

PENG, Roger D.; CAFFO, Brian; LEEK Jeffrey; **Executive Data Science: A Guide to Training and Managing the Best Data Scientists**, Leanpub, 2018.

PRESSMAN, Roger S.; MAXIM, Bruce R; **Software engineering: a practitioner's approach**. 8ª ed, São Paulo: McGraw Hill Brasil, 2015.

PROVOST, Foster; FAWCETT, Tom. **Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking**. Sebastopol: O’Reilly Media, 2013.

ROEVER, Leonardo, **Guia Prático de Revisão Sistemática e Metanálise**, 1ª ed, Rio de Janeiro, Thieme Brazil, 2020.

SKIENA, Steven S.; **The data science design manual**, Springer, 2017.

SOMMERVILLE, Ian; **Software Engineering**, 9ª ed, Addison-Wesley, 2011

THUSOO, Ashish; SARMA, Joydeep Sen; **Creating a data-driven culture for DataOps**. Sebastopol: O'Reilly Media, 2019.

TREWIN, Simon. **The DataOps Revolution: Delivering the Data-Driven Enterprise**. Auerbach Publications, 2021.

TSUI, Frank; KARAM, Orlando; BERNAL, Barbara. **Essentials of software engineering**. 5ª ed, Jones & Bartlett Learning, 2022.

WALLS, Mandi. **Building a DevOps culture**. “O’Reilly Media, Inc.”, 2013.

- [1] Discente do curso de Especialização em Gestão de Projetos e Negócios em Tecnologia da Informação, do Instituto Federal do Rio de Janeiro, Campus Engenheiro Paulo de Frontin, e-mail: br.alves.jessica@gmail.com
- [1]Discente do curso de Especialização em Gestão de Projetos e Negócios em Tecnologia da Informação, do Instituto Federal do Rio de Janeiro, Campus Engenheiro Paulo de Frontin, e-mail: ksthales@gmail.com
- [2] Discente do curso de Especialização em Gestão de Projetos e Negócios em Tecnologia da Informação, do Instituto Federal do Rio de Janeiro, Campus Engenheiro Paulo de Frontin, e-mail: wesleydesateles@gmail.com
- [3] Docente do curso de Especialização em Gestão de Projetos e Negócios em Tecnologia da Informação, do Instituto Federal do Rio de Janeiro, Campus Engenheiro Paulo de Frontin, e-mail: daniel.borges@ifrj.edu.br

[← Post anterior](#)

RevistaFT

A RevistaFT é uma **Revista Científica Eletrônica Multidisciplinar Indexada de Alto Impacto e Qualis “B2” em 2023**. Periodicidade mensal e de acesso livre. Leia gratuitamente todos os artigos e publique o seu também [clikando aqui](#).

Contato

Queremos te ouvir.
WhatsApp RJ: (21) 98159-7352
WhatsApp SP: (11) 98597-3405
e-Mail:
contato@revistaft.com.br
ISSN: 1678-0817
CNPJ: 48.728.404/0001-22

Conselho Editorial

Editores Fundadores:
Dr. Oston de Lacerda Mendes.
Dr. João Marcelo Gigliotti.
Editor Científico:
Dr. Oston de Lacerda Mendes
Orientadoras:
Dra. Hevellyn Andrade Monteiro
Dra. Chimene Kuhn Nobre



CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), fundação do Ministério da Educação (MEC), desempenha papel fundamental na expansão e consolidação da pós-graduação stricto sensu (mestrado e doutorado) em todos os estados da Federação.

Revisores:

Lista atualizada periodicamente em revistافت.com.br/expediente
Venha fazer parte de nosso time de revisores também!