

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA
E TECNOLOGIA DO AMAPÁ – IFAP
CÂMPUS MACAPÁ
CURSO SUPERIOR DE TECNOLOGIA EM REDES DE COMPUTADORES

HAILA BIANCA DOS SANTOS DE OLIVEIRA

DATAOPS: O NOVO PARADIGMA DE METODOLOGIA ÁGIL

MACAPÁ - AP

2020

HAILA BIANCA DOS SANTOS DE OLIVEIRA

DATAOPS: O NOVO PARADIGMA DE METODOLOGIA ÁGIL

Trabalho de Conclusão de Curso apresentado ao curso Superior de Tecnologia em Redes de Computadores, do Instituto Federal de Educação, Ciência e Tecnologia do Amapá – Ifap, como requisito avaliativo para obtenção de título de Tecnólogo em Redes de Computadores.

Orientador: Prof. Klenilmar Lopes Dias

MACAPÁ - AP

2020

Biblioteca Institucional - IFAP
Dados Internacionais de Catalogação na Publicação (CIP)

O048d Oliveira, Haila Bianca dos Santos de
 DataOps: O Novo Paradigma de Metodologia Ágil / Haila Bianca dos
 Santos de Oliveira - Macapá, 2020.
 39 f.

 Trabalho de Conclusão de Curso (Graduação) -- Instituto Federal de
 Educação, Ciência e Tecnologia do Amapá, Campus Macapá, Curso de
 Tecnologia em Redes de Computadores, 2020.

 Orientador: Dr. Klenilmar Lopes Dias.

 1. DataOps. 2. Pipeline de dados. 3. Fluxo de dados. I. Dias, Dr.
 Klenilmar Lopes , orient. II. Título.

HAILA BIANCA DOS SANTOS DE OLIVEIRA

DATAOPS: O NOVO PARADIGMA DE METODOLOGIA ÁGIL

Trabalho de Conclusão de Curso apresentado ao curso Superior de Tecnologia em Redes de Computadores, do Instituto Federal de Educação, Ciência e Tecnologia do Amapá – Ifap, como requisito avaliativo para obtenção de título de Tecnólogo em Redes de Computadores.

Orientador: Prof. Klenilmar Lopes Dias

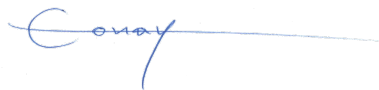
BANCA EXAMINADORA



Prof. Dr. Klenilmar Lopes Dias



Prof. Me. Klessis Lopes Dias



Prof. Esp. Eonay Barbo Gurjão

Aprovado(o) em: 18/12/2020

Nota: 9,83

AGRADECIMENTOS

Agradeço a Deus por me dar forças durante todos os momentos felizes e difíceis encontrados ao longo do curso.

Aos meus pais e irmãos, que me incentivaram nos momentos difíceis e apoiando em todas as decisões que tomei para fortalecer minha carreira acadêmica.

Ao meu professor orientador que me apoiou e me ajudou em todo o processo deste trabalho, dando todo auxílio e suporte que precisei, e aos meus professores pelos ensinamentos que permitiram o meu melhor desempenho para minha formação profissional.

“A perfeição é algo inatingível, mas se perseguirmos a perfeição podemos alcançar a excelência. Ser ágil é concentrar-se em fornecer a melhor coisa em um período de tempo definido.”

Vince Lombardi

RESUMO

DataOps é o novo paradigma dos dados, ele uma metodologia automatizada, orientada a processos utilizada por equipes analíticas e engenheiros de dados que agiliza a produção de alta qualidade de insights de dados com segurança e controle no fluxo de dados, já que o volume de dados cresce estrondosamente a cada dia provocando que empresas inovem em seus sistemas de dados, mas a maioria das vezes de forma incorreta, causando ineficiência em atender os requisitos de demanda do mercado. Essa metodologia tem a fusão com a metodologia Devops, que vai melhorar o desenvolvimento de software e das operações de TI como velocidade, a qualidade, a previsibilidade e a escala da engenharia e implantação de software com uma boa equipe com estratégias projetadas totalmente em governança priorizando sempre a segurança e a privacidade do cliente. O objetivo deste trabalho é informar bibliograficamente esse modelo de metodologia ágil revolucionária com o levantamento de informações para "O que ele é", quais são as etapas, requisitos de construção de um ambiente, e como é formação de uma equipe DataOps, abordando como funciona, suas ferramentas, vantagens, as sua diferenças com DevOps e contribuir com o incentivo para empresas e profissionais da área de Tecnologia da Informação na implementação desta prática apresentando cuidadosamente o DataOps como um novo modelo ágil de prática para as empresas entenderem a importância do uso da análise de dados na estratégia corporativa em forma de referencial teórico, também promovendo a procura por mais matérias desta pratica que são tão poucas no Brasil.

Palavras-chave: DataOps. Pipeline de dados. Fluxo de dados.

ABSTRACT

DataOps is the new data paradigm, it is an automated, process-oriented methodology used by analytical teams and data engineers that streamlines the production of high-quality data insights with security and control in the data flow, since the data volume it grows tremendously every day, causing companies to innovate in their data systems, but most of the time incorrectly, causing inefficiency in meeting market demand requirements. This methodology merges with the Devops methodology, which will improve software development and IT operations such as speed, quality, predictability and scale of engineering and software deployment with a good team with strategies designed entirely in governance, prioritizing always the customer's security and privacy. The objective of this work is to bibliographically inform this model of revolutionary agile methodology with the collection of information for "What it is", what are the steps, requirements for building an environment, and how is it to form a DataOps team, addressing how it works , its tools, advantages, its differences with DevOps and contribute to the incentive for companies and professionals in the area of Information Technology in the implementation of this practice, carefully presenting DataOps as a new agile practice model for companies to understand the importance of using analysis of data in the corporate strategy in the form of a theoretical framework, also promoting the search for more subjects of this practice that are so few in Brazil.

Keywords: DataOps. Data pipeline. Data flow.

LISTA DE FIGURAS

Figura 1 - Crescimento global do usuário da Internet. p.11

Figura 2 - Dispositivo global e crescimento da conexão. p.12

Figura 3 - DataOps combina os conceitos Ágil, DevOps e Controle de Processos Estatísticos e os aplica nos processos analíticos. p.15

Figura 4 - Ciclo de vida de DevOps como um loop infinito. p. 22

Figura 5 - DataOps no fluxo de inovação e valor. p. 22

Figura 6 - Fluxograma das ferramentas DataOps. p.30

Figura 7 - Fluxograma das vantagens DataOps. p.33

LISTA DE ABREVIATURAS E SIGLAS

APIs	Interface de programação de aplicações
DP	Aprendizado Profundo - Deep Learning
GDPR	Regulamento Geral sobre a Proteção de Dados
IA	Inteligência Artificial
M2M	Machine to Machine
ML	Aprendizado de Máquina - Machine Learning
TI	Tecnologia da Informação

SUMÁRIO

1	INTRODUÇÃO	11
2	DATAOPS: COMO CHEGAMOS AQUI ?	15
2.1	Como surgiu?	16
2.2	Princípios DataOps	17
2.3	Como o DataOps funciona ?	19
2.4	Fases DataOps	21
3	DATAOPS VS DEVOPS	23
3.1	Processos DataOps e DevOps	23
3.2	Diferenças DataOps vs DevOps	25
3.3	Segurança	26
4	AMBIENTES DATAOPS	28
4.1	Implementação	28
4.2	Ferramentas	31
4.3	Vantagens	34
5	CONCLUSÃO	37
	REFERÊNCIA BIBLIOGRÁFICA	38

1 INTRODUÇÃO

Com o aumento exponencial da colaboração, integração e aumento da complexidade das operações das organizações, surgiu a necessidade de agilizar e otimizar os fluxos de processos para entregar mais valor e qualidade às atividades oferecidas pelas organizações. Isto significa que as operações tradicionais de TI (Tecnologia da Informação) não estão conseguindo atender aos requisitos dessa nova demanda de empreendedorismo e necessidades de suprir os requisitos de mercado baseadas em TI.

Assim, com um conjunto de teorias, metodologias, processos, tecnologias e estruturas que transformam grandes quantidades de dados que, sozinhos, não significam muito, em informações essenciais para uma boa gestão. Ou seja, segundo Data Science Academy (2020) DataOps é uma metodologia automatizada, orientada a processos, usada por equipes analíticas e de dados, para melhorar a qualidade e reduzir o tempo de ciclo da análise de dados. Embora o DataOps tenha começado como um conjunto de práticas recomendadas, ele agora amadureceu e se tornou uma abordagem nova e independente da análise de dados. O DataOps se aplica a todo o ciclo de vida dos dados, da preparação dos dados aos relatórios, e reconhece a natureza interconectada da equipe de análise de dados e das operações de tecnologia da informação.

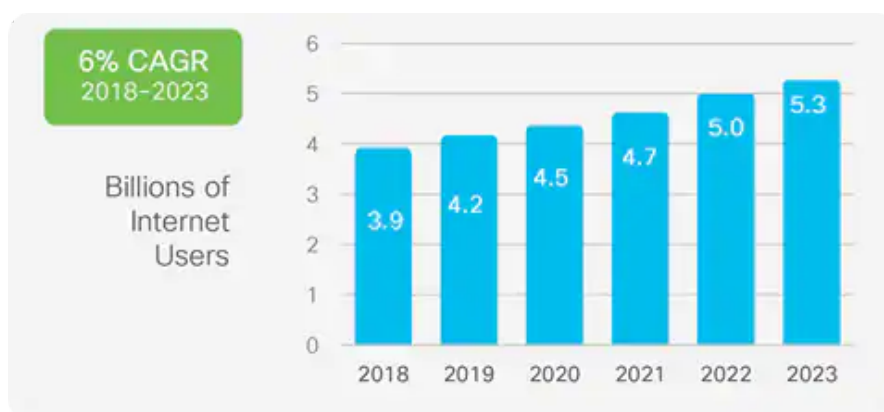
Com isso os dados são o principal ativo de uma organização, pois auxiliam na melhor tomada de decisão, análise, desempenho, podendo solucionar problemas, analisar o comportamento do consumidor e do mercado e proceder o mais rápido possível. Além disso, os dados são a espinha dorsal de muitas tecnologias de tendência, como ML (Aprendizado de Máquina - Machine Learning), DP (Aprendizado Profundo - Deep Learning) e IA (Inteligência Artificial). A crescente importância dos dados levou à aquisição e armazenamento de grandes volumes de dados, o que por sua vez levou a origem a áreas como Big Data, mineração de dados (Data Mining) e computação em nuvem (Cloud Computing). As operações iniciadas por engenheiros ou as operações realizadas alterando as fontes de dados causarão mudanças contínuas de dados, portanto, versões de dados e tecnologias de compartilhamento são necessárias.

O Paradigma DataOps surgiu com a mesma ideia do DevOps. O DevOps se concentra na entrega contínua, aproveitando os recursos de TI sob demanda e automatizando o teste e a implantação de softwares e aplicações. Essa fusão do desenvolvimento de software e das

operações de TI melhorou a velocidade, a qualidade, a previsibilidade e a escala da engenharia e implantação de software. Emprestando métodos do DevOps, o DataOps procura trazer esses mesmos aprimoramentos para a análise de dados, sempre prezando pela integração contínua e segurança na rede, GDPR (Regulamento Geral sobre a Proteção de Dados) é um bom exemplo, que adotou DataOps para gerenciamento de dados de próxima geração, e garantindo que os dados ainda sejam acessíveis e valiosos para que as organizações possam usar ativos de informação para impulsionar a inovação.

Com cada ano que passa, a taxa de crescimento de usuários na internet tem ocasionado no grande volume de dados nas redes de internet no planeta, a maioria das pessoas que estão ativas possuem celulares, principalmente pelas redes sociais, uma pesquisa realizada pela CISCO (Cisco Systems, Inc.) mostra um gráfico com uma pesquisa do aumento de usuários que pode chegar até 2023.

Figura 1 - Crescimento global do usuário da Internet



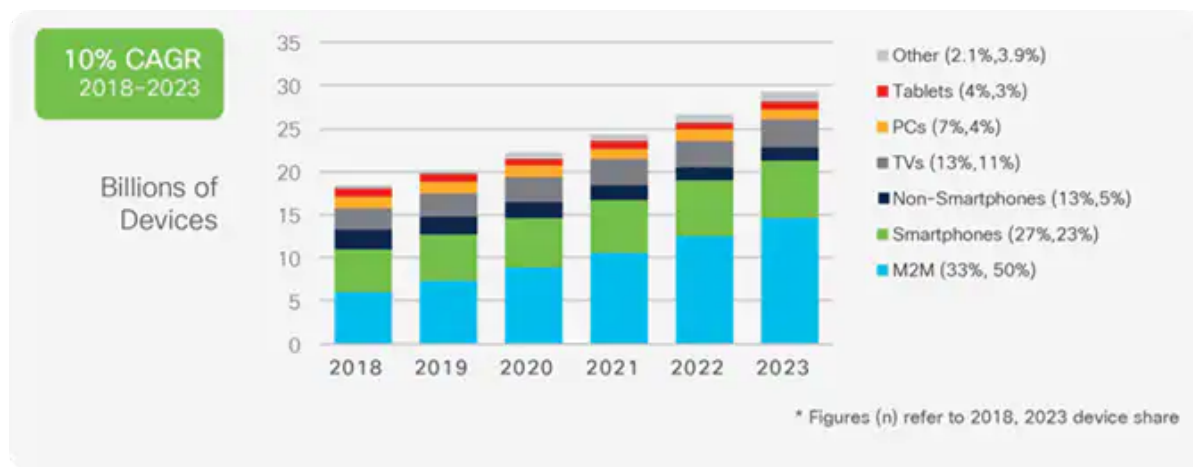
Fonte: Relatório Anual da Internet da Cisco, 2018-2023

CISCO (2020) Globalmente, o número total de usuários da Internet está projetado para crescer de 3,9 bilhões em 2018 para 5,3 bilhões em 2023 a um CAGR de 6%. Em termos de população, isso representa 51% da população global em 2018 e 66% da penetração da população global em 2023 (Figura 1). Porém não é só o volume de dados que está crescendo, também dispositivos e conexões principalmente por domicílio e per capita.

Globalmente, os dispositivos e conexões estão tendo a tendência de acelerar o aumento, mais rápido 10% CAGR do que a população 1,0% CAGR e os usuários da Internet 6% CAGR. A cada ano, vários novos dispositivos em diferentes formatos com recursos e inteligência aumentados são introduzidos e adotados no mercado. Em 2023, as conexões

M2M (Machine to Machine) representarão metade ou 50 por cento do total de dispositivos e conexões (Figura 2).

Figura 2 - Dispositivo global e crescimento da conexão



Fonte: Relatório Anual da Internet da Cisco, 2018–2023

Com o volume dados subindo a cada dia, é inevitável que empresas corram contra o tempo para inovar seus sistemas de rede com equipamentos e softwares de última linha, porém isso não é o bastante, sem uma equipe e um projeto não será definido qual é o objetivo da gestão de dados e terá falhas de integração entre as áreas.

A busca pela inovação provoca rapidez para que empresas utilizem dados disponíveis para se impulsionar no mercado, o que faz os dados menos seguros serem distribuídos de forma incorreta sobrecarregando equipes de não produção causando vulnerabilidade a ataques cibernéticos.

Com DataOps o modo de trabalho será inter-funcional, no contexto da aquisição, armazenamento, processamento, monitoramento de qualidade, performance, análise preditiva, escalabilidade de backups, melhoria e entrega de informações ao usuário final, atuando com um projeto de governança para o negócio da empresa.

O movimento DataOps tem ganhado força quando se fala de negócios inteligentes e segurança, suas técnicas de automatização são essenciais em empresas de tecnologia baseada em dados, e tem impulsionado o crescimento de negócios de médio e grande porte, desta forma não é mais possível abrir mão de tecnologia e metodologias de inovação e automação quando se fala de qualquer negócio que utilize uma infraestrutura de rede.

Portanto, a proposta central deste trabalho é apresentar o processo de automação de

DataOps e apresentar cuidadosamente o DataOps como um novo modelo de prática para as empresas entenderem a importância do uso da análise de dados na estratégia corporativa em forma de referencial teórico.

Com objetivo de levantamento de informações para o conhecimento do “o que é” e quais são as etapas, requisitos de construção de um ambiente, e como é formação de uma equipe DataOps, abordando suas vantagens na agilidade e segurança de sistema de rede de uma empresa.

2 DATAOPS: COMO CHEGAMOS AQUI ?

O que é DataOps ? o conceito não é tão simples assim, principalmente porque DataOps surgiu da comunidade de programadores e até hoje não existe uma empresa ou instituição por trás desse movimento DataOps de potências do mercado atual, e mesmo que cada um ponha o seu ponto de vista estamos falando da mesma coisa, ou seja, DataOps é a combinação de filosofias culturais práticas e ferramentas para que existam alinhamento e colaboração em toda a empresa, desde a TI e os especialistas em dados até os consumidores de dados para entrega rápida de serviços com alto grau de qualidade.

Segundo IBM (2020) O DataOps (operações de dados) é a orquestração de pessoas, processos e tecnologia para entregar dados confiáveis e prontos para os negócios aos cidadãos de dados, às operações, aos aplicativos e à inteligência artificial (IA) rapidamente.

Ops em DataOps é um lembrete importante, quais dados devemos ir além do gerenciamento e análise de dados é a forma como operamos o fornecimento de dados e a saída do resultado. Na verdade, transferência de dados e pipeline de dados são apenas componentes de grandes aplicativos de dados envolvendo várias equipes. Portanto, DataOps deve incluir casos de uso de aplicativo, cada equipe envolvida no ciclo de vida dos dados irá pegar os dados que você consome e o resultado final e muitos benefícios.

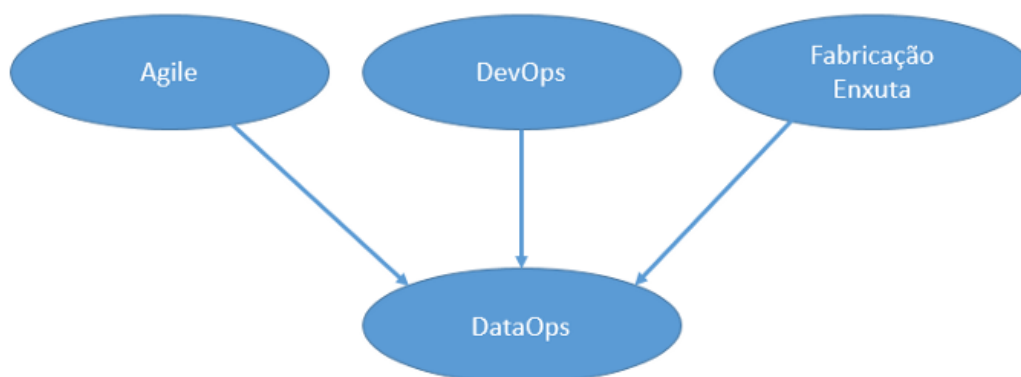
Esse paradigma, que combina os conceitos Ágil, DevOps e Fabricação Enxuta (Controle de Processos Estatísticos), busca simplificar processos e acelerar a entrega de dados, garantindo a qualidade e precisão das informações em projetos de Machine Learning e Data Analytics.

Bases do DataOps (Figura 2):

- **Agile:** otimiza fluxos de trabalho, melhora a produtividade de projetos e eleva significativamente as perspectivas de sucesso do seu negócio.
- **DevOps:** propõe novos pensamentos sobre o trabalho para a valorização da diversidade de atividades e profissionais envolvidos e atitudes colaborativas. É um processo que torna possível o desenvolvimento ágil de aplicações em um modelo de gestão de infraestrutura definido sob regras rígidas e burocráticas.
- **Produção Enxuta:** visa a eliminação progressiva do desperdício, pelo fluxo contínuo com que os processos produtivos ocorrem, pela produção segundo a demanda do

cliente, no tempo e na quantidade por este estabelecidos e, por fim, pela relação próxima e de parceria com fornecedores.

Figura 3 - DataOps combina os conceitos Ágil, DevOps e Controle de Processos Estatísticos e os aplica nos processos analíticos.



Fonte: O que é DataOps: Organizando o futuro do Data Science para negócios (2019)

No desenvolvimento de um sistema, as práticas DevOps e Agile se encaixam como uma luva para levantar os requisitos, efetuar o desenvolvimento, testar e entregar um produto de maneira rápida e contínua, mas quando tratamos de sistemas analíticos, é necessário também levar em consideração a arquitetura do fluxo de dados (data pipelines), pois os dados estarão continuamente entrando na alimentação de recursos utilizados pelo sistema, assim como alimentando o próprio sistema. Esse fluxo de dados é o lado “operacional” dos sistemas analíticos. É justamente aqui que se inicia a filosofia DataOps, que começa trazendo boas práticas que buscam diluir as barreiras e complicações que existem entre as áreas de desenvolvimento e operações analíticas. (LOPES, 2019)

2.1 Como surgiu?

O termo DataOps surgiu pela primeira vez em junho de 2014, em um post do blog IBM Big Data & Analytics Hub com o título “3 reasons why DataOps is essential for big data success” (3 razões pelas quais DataOps é essencial para o sucesso de big data, na tradução literal), de Lenny Liebmann.

Segundo Liebmann (2014) Antigamente, os desenvolvedores escreviam o código do aplicativo e simplesmente “jogavam fora” para as operações de TI, que então tinham que garantir que esses aplicativos funcionassem bem no ambiente de produção. Essa abordagem sempre foi menos do que ideal, mas se tornou insustentável à medida que o negócio começou a depender cada vez mais de muitos códigos novos sendo lançados em produção rapidamente e com alto grau de confiança. Portanto, as organizações de TI agora estão adotando um conjunto de práticas recomendadas conhecidas como DevOps, que melhoram a coordenação entre o desenvolvimento e as operações.

O autor definiu, então, DataOps como sendo o conjunto de práticas que melhoram e facilitam a coordenação de Ciência de Dados e Operações, o objetivo geral de acelerar a entrega de dados, garantindo a qualidade e a acuracidade das informações e assim fornecendo uma organização data-driven, que toma decisões baseadas em dados, que precisam de respostas rápidas e precisas para atender ao timing do negócio.

O autor cita três razões para o diferencial do DataOps :

- **Razão 1:** a velocidade conta. As oportunidades de negócios geralmente têm vida útil curta. Em muitos casos, podem ser tão fugazes quanto uma visita a um site ou um telefonema. Portanto, para todos os efeitos práticos, resultados lentos podem não ser resultados.
- **Razão 2:** a nuvem não é uma panaceia. É incrível que muita capacidade de processamento relativamente barata esteja disponível sob demanda de provedores de serviços em nuvem. Mas nem todo problema de desempenho de big data pode ser resolvido ativando um monte de VMs Hadoop.
- **Razão 3:** as cargas de trabalho de Big Data são diversas. Big data não é apenas uma coisa. Um dia, é uma análise preditiva. No dia seguinte, é a veiculação de dados móveis. No dia seguinte, é o processamento da transação.

2.2 Princípios DataOps

DataOps é o contexto da aquisição, armazenamento, processamento, monitoramento de qualidade e performance, melhoria, e entrega de informações ao usuário final, de forma contínua e confiável, portanto, uma disciplina importantíssima para qualquer organização de TI que deseja prosperar no mundo da inteligência de negócios nos tempos de hoje é uma necessidade competitiva. De acordo com o manifesto de DataOps, os princípios devem ser:

- **Satisfazer continuamente o seu cliente:** A prioridade é satisfazer o cliente através da entrega antecipada e contínua de informações analíticas valiosas que podem ser de 2 minutos até duas semanas.
- **Valorizar o trabalho analítico:** A principal medida do desempenho da análise de dados é o grau em que as análises são entregues, incorporando dados precisos ao invés de dados e sistemas robustos.
- **Abraçar a mudança:** Acolher as crescentes necessidades do cliente e, de fato, abraçar essas necessidades a fim de gerar vantagem competitiva. Acreditar que o método de comunicação mais eficiente, eficaz e ágil com os clientes é uma conversa cara a cara.
- **Trabalhar em equipe:** As equipes sempre terão uma variedade de papéis, habilidades, ferramentas favoritas e títulos.
- **Interações diárias:** Clientes, equipes de análise de dados e operações devem trabalhar juntas durante todo o projeto.
- **Auto-organização:** A melhor visão analítica, algoritmos, arquiteturas, requisitos e projetos emergem de equipes auto-organizadas.
- **Reduzir o heroísmo:** À medida que o ritmo e a amplitude da necessidade de insights analíticos aumentam, as equipes devem se esforçar para reduzir o heroísmo e criar equipes e processos sustentáveis e escaláveis.
- **Reflexão:** Equipes de análise de dados devem aperfeiçoar seu desempenho nas operações através de uma autorreflexão, em intervalos regulares, sobre os feedbacks fornecidos por seus clientes, por eles próprios e pelas estatísticas operacionais.
- **Os códigos:** As equipes usam uma variedade de ferramentas para acessar, integrar, modelar e visualizar dados. Fundamentalmente, cada uma dessas ferramentas gera códigos e configurações que descrevem as ações tomadas sobre dados para fornecer informações.
- **Orquestração:** Orquestração de dados, ferramentas, códigos, ambientes e equipes de trabalho são fatores-chave para o sucesso dos projetos de análise de dados.
- **Faça tudo ser reproduzível:** São necessários resultados reproduzíveis e, portanto, versionar tudo: dados, configurações de hardware e software, código e configurações específicas de cada ferramenta utilizada.
- **Ambientes descartáveis:** É importante minimizar o custo para os membros das equipes de análise de dados fazerem experimentações, proporcionando-lhes facilidade

de criar ambientes técnicos descartáveis, isolados e seguros que reflitam o ambiente de produção.

- **Simplicidade:** A atenção contínua à excelência técnica e ao bom design aumenta a agilidade; bem como a simplicidade, a arte de maximizar a quantidade de trabalho não feito é essencial.
- **Análise de dados é manufatura:** Os pipelines de análise de dados são análogos às linhas de fabricação enxuta. Um conceito fundamental de DataOps é o foco no pensamento processual destinado a alcançar eficiência contínua na construção de insights analíticos.
- **A qualidade é primordial:** Os pipelines de análise de dados devem ser construídos com uma fundação capaz de detectar automaticamente erros no código, configurações e dados, e devem fornecer feedback contínuo aos operadores para evitar erros.
- **Monitorar a qualidade e o desempenho:** O objetivo é ter medidas de desempenho e qualidade que sejam monitoradas continuamente para detectar variações inesperadas e gerar estatísticas operacionais.
- **Reutilização:** Um aspecto fundamental da eficiência na fabricação de insights analíticos é evitar a repetição do trabalho anterior pelo indivíduo ou pelo time.
- **Melhorar os tempos dos ciclos:** Devemos nos esforçar para minimizar o tempo e o esforço para transformar a necessidade de um cliente em uma ideia analítica, criá-la em desenvolvimento, liberá-la como um processo de produção repetível e, finalmente, refatorar e reutilizar esse produto.

Os princípios de DataOps buscam estimular a aplicação da metodologia Ágil em qualquer área que envolva desenvolvimento de sistemas. O foco é sempre diminuir as distâncias que precisam ser percorridas para a entrega do produto e quebrar as barreiras entre as diferentes áreas no desenvolvimento e manutenção do produto.

2.3 Como o DataOps funciona ?

A cada dia empresas estão migrando para machine learning com uma vasta lista de serviços, e Dataops chega orientando e apoiando as necessidades de aprendizado de máquina de ponta a ponta.

Mas Dataops não se limita a só machine learning, ted dunning fala em um artigo no site de computerworld (2018) que Dataops é útil para qualquer trabalho orientado a dados,

tornando mais fácil aproveitar os benefícios oferecidos pela construção de um tecido de dados global. Observando também que Dataops se encaixa perfeitamente com arquiteturas dos microsserviços.

A abordagem DevOps dentro do DataOps faz a junção das equipes em desenvolvimento e operações para alinhar o desenvolvimento de software com os objetivos de negócios, encurtar os ciclos de desenvolvimento e aumentar a frequência de implantação. Enfatiza equipes multifuncionais como operações, engenharia de software, arquitetura e planejamento e gerenciamento de produtos. E com elas exibe as funções de ciência de dados e engenharia de dados, com o objetivo de aumentar a colaboração e a comunicação entre desenvolvedores, profissionais de operações e especialistas em dados.

A coisa mais importante a fazer aqui é não ficar com a organização mais tradicional da Ivory Tower onde os cientistas de dados vivem separados das equipes de desenvolvimento, o passo mais importante que você pode tomar é incorporar cientistas de dados em uma equipe DevOps. Quando eles moram no mesmo quarto, comem as mesmas refeições, ouvem as mesmas queixas, eles se alinham. (computerworld, 2018)

Aqui estão as atividades principais do DataOps (DATA SCIENCE ACADEMY, 2020):

- Estabelecer medições de progresso e desempenho em todas as etapas do fluxo de dados. Sempre que possível, avaliar os tempos do ciclo de fluxo de dados.
- Definir regras para uma camada semântica abstraída. Garantir que todos estejam “falando o mesmo idioma” e concordem com o que os dados (e metadados) são e não são.
- Incluir loops de feedback humano orientados para a melhoria contínua. Os consumidores devem poder confiar nos dados, e isso só pode vir com validação incremental.
- Automatizar o maior número possível de estágios do fluxo de dados, incluindo BI, Data Science e Análise.

- Usar informações de desempenho, identificar gargalos e otimizá-los. Isso pode exigir investimento em hardware básico ou automação de uma etapa de Ciência de Dados fornecida anteriormente por humanos no processo.
- Estabelecer disciplina de governança, com foco especial no controle de dados bidirecional, propriedade dos dados, transparência e rastreamento abrangente da linhagem de dados em todo o fluxo de trabalho.
- Processo de design para crescimento e extensibilidade. O modelo de fluxo de dados deve ser projetado para acomodar volume e variedade de dados. Garantir que as tecnologias facilitadoras tenham um preço acessível para escalar com o crescimento de dados corporativos.

2.4 Fases DataOps

No DataOps, há vários estágios que constituem o processo. Começa com a aquisição de dados, passando por armazenamento e monitoramento de qualidade e desempenho, e então conduz uma série de testes, análises preditivas e melhorias em um processo de entrega contínuo e cuidadosamente organizado. Resumindo, é assim:

Análise > Desenvolvimento > Orquestração > Teste > Entrega > Orquestração > Administração

A primeira etapa é a análise dos dados, que percorre o processo de desenvolvimento até chegar à parte do processo de negócio, uma das etapas mais importantes do processo. Primeiro, ele organiza os dados, trata as exceções e as atribui ao próximo fluxo.

A partir daí, teste até a entrega. Em seguida, ela voltou ao segundo momento da orquestração, o que permitiu à mesma organização monitorar o controle de erros durante os testes posteriores. Por fim, a última etapa é a gestão, cujo objetivo é otimizar o uso dos recursos.

O objetivo do DataOps para ciência de dados é transformar dados brutos em dados úteis aos produtos de ciência de dados que podem ser passados de forma rápida, escalável e Processo repetível. Os produtos de ciência de dados integram ciência de dados às operações Serviço ou produto. Várias vezes ao dia, somos todos clientes de produtos de dados como Netflix e Google Maps.

Os produtos de ciência de dados estão em produção contínua sob monitoramento contínuo, a iteração baseada em experimentação e feedback pode trazer melhorias. Está tendo um proprietário, reproduzível e resolve o objetivo final. O usuário e a máquina podem interagir usando produtos de ciência de dados de várias maneiras, como APIs (Interface de programação de aplicações), visualizações e até mesmo web ou Interface de aplicativo móvel.

DataOps é uma solução para muitos dos desafios enfrentados pela ciência de dados e análise, uma vez que você percebe que dados brutos devem ser tratados como um processo de linha de montagem ponta a ponta que requer altos níveis de colaboração, automação e melhoria contínua.

3 DATAOPS VS DEVOPS

No mundo agile ocorre um equívoco com DataOps que ele é um “DevOps dos dados”, o que é totalmente errado do que é o conceito verdadeiro como dito no capítulo 1, Dataops é a análise de dados que pode alcançar muito mais o que o desenvolvimento de software alcançou com DevOps.

O DevOps tem como objetivo na integração e entrega contínua de software, aproveitando recursos de TI sob demanda e infraestrutura como código e automatizando a integração, teste e implantação de código. Essa fusão de desenvolvimento de software e operações de TI reduz o tempo de implantação, tempo de colocação no mercado, defeitos e o tempo necessário para resolver problemas. DevOps merece muito crédito pelo sucesso de empresas como Google, Amazon e Microsoft, e lançam software muitas vezes por dia que melhora a qualidade e o tempo de ciclo dos lançamentos de código.

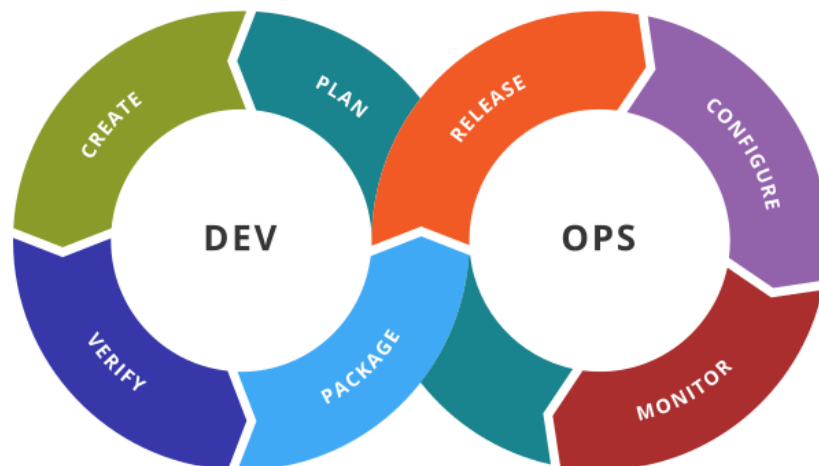
Porém a implantação de análise de dados é muito mais complexa, DataOps está empenhada em reduzir o tempo de ciclo do início ao fim da análise de dados, da origem da ideia à criação de tabelas, gráficos e modelos de criação de valor. O ciclo de vida dos dados depende de pessoas. Para que o DataOps seja eficaz, ele deve gerenciar a colaboração e a inovação. Para isso, o DataOps introduz o desenvolvimento ágil na análise de dados, para que as equipes de dados e os usuários possam trabalhar juntos com mais eficácia.

DevOps propõe para a entrega de software, e o DataOps pela entrega de dados, e o que torna eles semelhantes é a metodologia ágil e mesmo assim diferentes com suas práticas e ambientes.

3.1 Processos DataOps e DevOps

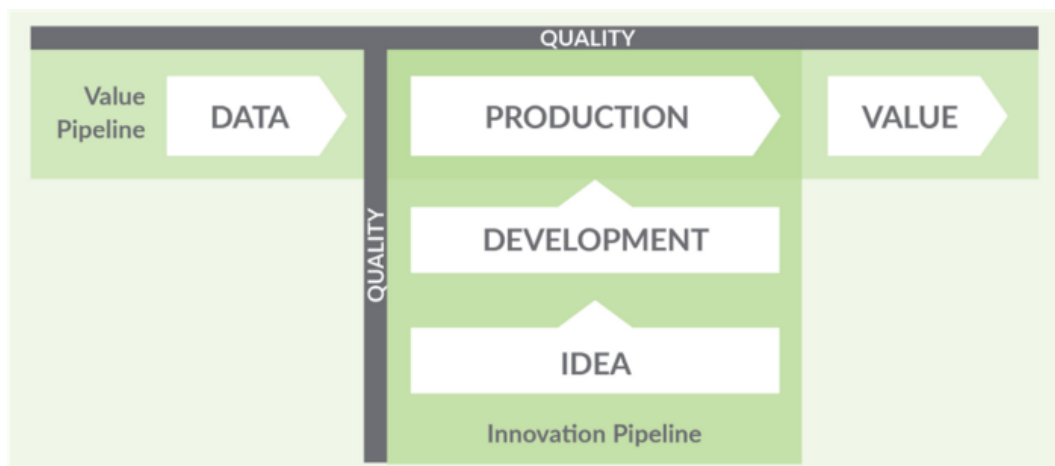
DevOps é frequentemente descrito como um loop infinito, esse formato serve para representar a atividade constante que a metodologia exige. Enquanto DataOps é ilustrado como canais de valor e inovação que se cruzam.

Figura 4 - Ciclo de vida de DevOps como um loop infinito.



Fonte: DataOps is NOT Just DevOps for Data (2018)

Figura 5 - DataOps no fluxo de inovação e valor.



Fonte: DataOps is NOT Just DevOps for Data (2018)

Os engenheiros de sistema tendem a se aprofundar na diversidade da tecnologia e procuram todas as oportunidades de experimentar novas ferramentas e linguagens de programação. Cientistas e analistas de dados têm visões diferentes sobre o processo de desenvolvimento. Para eles, linguagens de programação e sistemas são ferramentas intermediárias de processamento e manipulação de dados e criação de seus modelos. Isso significa que em um ambiente DevOps, a cultura tende a amar a complexidade da tecnologia, enquanto em DataOps, a pesquisa é simplificar o uso da tecnologia no processo de criação, o que tende a desviar o foco de linguagens robustas e sistemas de alto desempenho. O desempenho de sistemas e ferramentas pode ajudar a simplificar o processo de modelagem e torná-lo mais fácil de gerenciar.

DataOps utiliza técnicas DevOps, além de adaptá-las para o processo de desenvolvimento analítico, como mostra o ciclo de vida da de DevOps e DataOps (Figura 3) e (Figura 4):

Fases do processo DevOps:

Desenvolvimento > Construção > Teste > Entrega > Administração

Fases do processo DataOps:

Análise > Desenvolvimento > Orquestração > Teste > Entrega > Orquestração > Administração

No DataOps o fluxo de valor e inovação funciona com "Pipeline de inovação" é o processo de introdução de novas ideias analíticas no pipeline de valor. O pipeline de inovação é conceitualmente semelhante ao processo de desenvolvimento DevOps, mas após observação cuidadosa, existem vários fatores que tornam o processo de desenvolvimento DataOps mais desafiador do que DevOps.

Tendo dois passos de orquestração sendo a primeira fase responsável por lidar com o processo conhecido como “fábrica dos dados”, que consiste num fluxo de processos com muitos passos, onde um “orquestrador” organiza a alimentação dos muitos fluxos de dados, monitora-os e lida com exceções. Já a segunda fase efetua a mesma organização, monitoramento e controle de erros dos processos nos testes pós produção.

LOPEZ (2019) A dualidade de fases de orquestração (orquestração dos dados e orquestração dos códigos) serve para dar ênfase no controle de qualidade que é necessário para um sistema de fluxo intenso de dados.

3.2 Diferenças DataOps vs DevOps

Tabela 1: DataOps vs DevOps

DataOps	DevOps
O ciclo de vida compartilha propriedades interativas, mas com uma diferença importante: 1- A fábrica de dados;	Seu ciclo de vida geralmente é ilustrado em um símbolo infinito : O final do ciclo retorna ao início e o processo intera indefinidamente.

2- Controle sobre como a fabrica de dados é atualizada.	
O processo consiste em: Gerenciamento de do SandBox Desenvolver , Orquestrar, Testar, Implementar, Orquestrar, Monitorar.	O processo consiste em: Desenvolver, Construir, Testar, Implementar, Rodar.
Conjunto de Práticas e Tecnologias que automatiza o ciclo de vida de ponta a ponta da análise de dados para melhorar a produtividade e agilidade .	Integra a área de desenvolvimento e operações.
Orienta-se para a cultura de dados.	Mudança cultural do departamento de TI.
Traz as sensibilidades do DevOps para o gerenciamento e a integração de dados.	A solução envolve a automatização da governança de TI e é ainda mais efetiva com aplicação do desenvolvimento agile.
Perfeita integração de novas análises à produção.	A equipe de operações recebe um Feedback contínuo de quem está programado: Integração entre diferentes times.
Gerenciamento eficiente, automação e confiabilidade / segurança no que tange a dados.	Otimiza os resultados da TI.

Fonte: Autor

3.3 Segurança

Os anos passam, e metodologias como DevOps impõe ainda mais demanda aos dados e, necessitando pelo grande volume dados que aumenta em grande velocidade a cada dia e a inovação traz produtos mais complexos e exigem mais recursos para atender a demanda por experiências personalizadas ao cliente, exigindo uma estratégia de dados sofisticados. Dados e acesso a ele são uma vantagem competitiva no mercado.

Porém o atrito de dados é um grande problema, é quando ocorre as restrições de dados impedem as pessoas de atender às crescentes demandas do negócio. (CySec Bot, 2018) Esse

atrito limita o número de ambientes de desenvolvimento, restringe a velocidade na qual o teste pode ser feito e impacta diretamente a capacidade de uma equipe de DevOps de obter integração e entrega contínuas. O que aumenta a complexidade e custo com crescentes preocupações de segurança e privacidade.

As empresas se acostumaram com processos em nuvem e DevOps, e entrando em guerra com os dados, os forçando a se encaixar com prática, o que faz os famosos “minutos” em “eternidade” causando congestionamento no sistema de rede.

(VECTOR ITC, 2017) Mas os dados não se parecem com cálculos. O tratamento de dados é caro para manter, cheio de informações confidenciais, difícil de copiar, difícil de rastrear ao longo do tempo, e lento para entregar a computadores que precisam dele.

Com DataOps, a plataforma de dados será dinâmica, que garante a segurança dos dados confidenciais e fornece os dados certos às pessoas certas no momento certo, para que os operadores e consumidores de dados possam trabalhar como um todo. (CySec Bot, 2018) O DataOps ajudará as organizações em todo o mundo a combater as dificuldades que surgirão, como GDPR, segurança de dados instável, a mudança para nuvens múltiplas e para ajudar o DevOps a alcançar seu verdadeiro potencial.

4 AMBIENTES DATAOPS

Em toda empresa de sucesso deve ter dados de alta qualidade no projeto analítico do sistema de ponta a ponta. Um bom projeto de governança é a grande solução para gerar valor no negócio e fornecer insights em tempo hábil. Porém, também é responsabilidade dele passar pelas etapas de processo de entrega de dados, requisitos e processos em rápida evolução, o que faz a equipe de planejamento reduzir os métodos tradicionais e entregando dados sem fundamento. Alguns roadblocks comuns incluem (GUPTA, 2019) :

- As equipes passam mais tempo identificando problemas de pipeline de dados e inconsistência de código devido ao código mais antigo ou conexão incorreta e informações de metadados, infraestrutura ou desafios relacionados a operações, ou resolvendo dependências técnicas entre as partes interessadas em comparação com o tempo gasto com foco na entrega de dados
- Os processos manuais levam a longos tempos de resposta, erros frequentes, dados inconsistentes e baixa repetibilidade, necessários para dar suporte a várias equipes continuamente.
- Processos isolados originados de economias sob demanda estão levando a dados inutilizáveis ou resultados imprevisíveis

No ambiente DataOps essas etapas (entrega de dados, requisitos e processos em rápida evolução) serão aprofundados entre as equipes para melhor harmonia com negócio e roadblocks irão se desbloquear gerando valor.

4.1 Implementação

DataOps não é uma prática privada, qualquer profissional de TI pode adotar este paradigma para seu sistema de dados. Para se ter uma boa implementação, etapas devem ser seguidas corretamente para alcançar todo o potencial dos dados. Primeiramente, avalie e ajuste seu portfólio e seus processos de tecnologia para remover a redundância e consolidar o controle em suas equipes.

Em seguida, reúna as equipes para estimular o compartilhamento e reduzir as inconsistências que impedem a colaboração. Em seguida, integre as práticas de DataOps em todas as equipes e pipelines de dados. Esse estágio geralmente é difícil porque a colaboração requer que seus funcionários apliquem processos desconhecidos e equipes confiáveis que nunca trabalharam juntas.

Após alinhar as equipes, você precisará automatizar o processo. A automação torna os pipelines de dados e a manipulação de dados mais eficiente. A última etapa é autorizar os usuários de dados a usarem eles próprios todos os recursos disponíveis. Os dados são rapidamente transformados em informações e percepções para revelar todo o potencial de seus DataOps, que agora está claro para toda a organização. Lembre-se de verificar todos os componentes de manipulação de dados e medir cuidadosamente seus processos para continuar a aprimorá-los, ajustá-los e atualizá-los para manter sempre as informações relevantes.

Segundo LOPES (2019) É necessário informar que não há uma receita para a implantação de uma filosofia. Os conceitos de DataOps buscam trazer a simplificação dos sistemas analíticos, o controle de “silos de dados”, a união entre o desenvolvimento e a operação e a fabricação enxuta de analíticos. Cada ambiente irá precisar de processos diferentes para implantação dessas técnicas e conceitos.

Porém, podemos definir alguns passos fundamentais que podem incrementalmente trazer o foco para a implantação dessas idéias no ambiente corporativo:

- **Testes autônomos nos fluxos de dados** : Com um foco maior no controle de qualidade dos dados, é importante implementar uma fase de testes para manter o controle de erros e fazer uma manutenção, observando erros e mantendo a exatidão dos dados.
- **Ferramentas de versionamento de código** : Ao utilizar ferramentas de versionamento, é possível manter a organização e estimular a automação da integração e entrega de códigos, inserindo os conceitos da filosofia DevOps durante esse processo.
- **Múltiplos ambientes de trabalho** : É consideravelmente mais simples colocar o sistema de promoção de ideias da filosofia DataOps em prática com o uso de ferramentas como containers, computadores e ambientes virtuais.
- **Reciclagem e unificação**: Evite a divisão de várias fórmulas e códigos diferentes para as mesmas medidas dentro da empresa. A padronização dos processos de modelagem,

previsões e visualizações é via de regra, para que haja uma evolução natural da organização e de processos autônomos.

- **Parametrização de processos:** Em determinadas situações, um processo analítico pode ser usado em várias condições diferentes. A reciclagem e agilidade no desenvolvimento depende diretamente da parametrização dos fluxos de dados, para que essas alterações sejam permitidas.
- **Armazenamentos simples:** É indicado fugir da utilização de sistemas de armazenamento muito complexos em relação à manutenção e gerenciamento. O armazenamento de dados deve ter seu foco em sistemas que possibilitam aos utilizadores focar na coleta e tratamento dos dados. Um exemplo disso são os sistemas de cloud computing.

Com cada passo sendo feito, irá gerar um fluxo de velocidade no sistema de dados, já que pequenos lotes de dados são mais rápidos de serem entregues.

“Pequenos pedaços completos de trabalho criam uma velocidade confiável e repetível e ajudam a determinar a capacidade da equipe. Em vez de muito trabalho parcialmente concluído, crie muitas unidades menores de trabalho completo. Para pipelines de dados, em vez de processar grandes lotes de dados, mova para mini ou micro lotes ou até mesmo streaming de dados usando uma arquitetura que responde aos dados assim que eles chegam”. (ATWAL,2019, p. 74)

DataOps pode acelerar a capacidade das equipes de análise de dados de criar e publicar novas análises para os usuários. Requer uma mentalidade ágil e também deve ser suportado por uma plataforma automatizada que incorpora ferramentas existentes em um pipeline de desenvolvimento DataOps (DATA KITCHEN, 2017).

O que a equipe DataOps tem a oferecer é de extrema importância, a prática DataOps utiliza vários tipos de ferramentas e métodos de várias áreas de TI para o perfeito funcionamento da própria. A equipe deve ser formada de um pouco de tudo, e se comunicarem e entenderem a necessidade para assim resolverem problemas.

4.2 Ferramentas

O DataOps não está vinculado a uma tecnologia, arquitetura, ferramenta, linguagem ou estrutura específica. As ferramentas que suportam o DataOps promovem colaboração, orquestração, qualidade, segurança, acesso e facilidade de uso.

Esse paradigma tem uma ampla gama de tecnologias e processos que facilitam para para sua equipe de TI, incluindo a tecnologia de gerenciamento de dados (catálogos de dados, virtualização de dados, pipelines de dados, gerenciamento de modelos de inteligência artificial), a tecnologia para controle de versões, a automação de testes, a automação de implantação, o gerenciamento de versões e a orquestração de tempo de execução ou, até mesmo, a colaboração. A automação de testes e implantações usa a IA e o ML para auxiliar nos processos e fluxos de trabalho, ajudando a evitar a configuração manual.

As ferramentas para suportar direta e indiretamente as necessidades de DataOps podem ser divididas em cinco etapas, aproveitando as ferramentas analíticas existentes junto com os componentes da cadeia de ferramentas destinados a abordar o gerenciamento de controle de origem, gerenciamento de processos e comunicação eficiente entre grupos para fornecer um pipeline de dados confiável.(GUPTA, 2019)

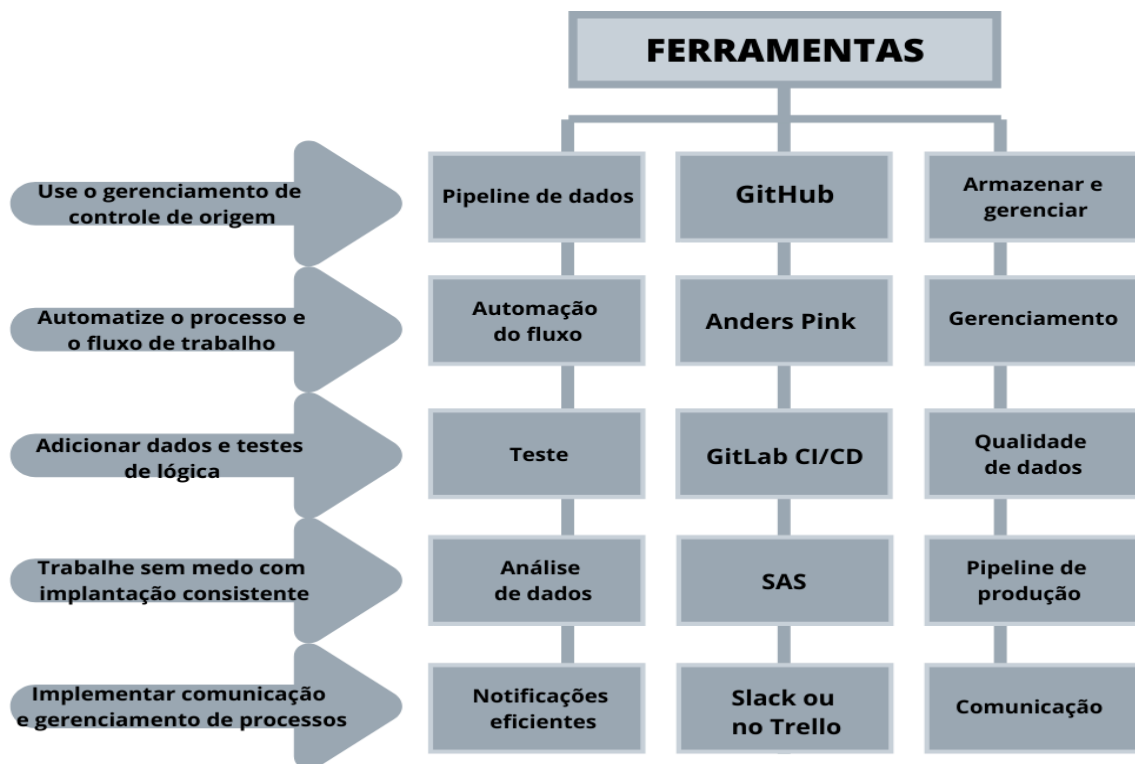
- **Use o gerenciamento de controle de origem:** um pipeline de dados nada mais é do que o código-fonte responsável por converter o conteúdo bruto em informações úteis. Podemos automatizar o pipeline de dados de ponta a ponta, produzindo um código-fonte que pode ser consumido de forma reproduzível. Uma ferramenta de controle de revisão (como o GitHub) ajuda a armazenar e gerenciar todas as alterações no código e na configuração para minimizar a implantação inconsistente.
- **Automatize o processo e o fluxo de trabalho do DataOps :** para que a metodologia DataOps seja bem-sucedida, a automação é a chave e requer um pipeline de dados projetado com flexibilidade de tempo de execução. Os principais requisitos para conseguir isso são serviços de curadoria de dados automatizados (como o Anders Pink), gerenciamento de metadados , governança de dados, gerenciamento de dados mestre e interação de autoatendimento.
- **Adicionar dados e testes de lógica:** para ter certeza de que o pipeline de dados está funcionando corretamente, o teste de entradas, saídas e lógica de negócios deve ser aplicado (GitLab CI/CD). Em cada estágio, o pipeline de dados é testado quanto à

precisão ou desvio potencial junto com erros ou avisos antes de serem liberados para ter uma qualidade de dados consistente.

- **Trabalhe sem medo com implantação consistente:** os profissionais de análise de dados temem a perspectiva de implantar mudanças que quebrem o pipeline de dados atual. Isso pode ser resolvido com dois fluxos de trabalho principais, que posteriormente se integram na produção. Primeiro, o pipeline de valor cria valor contínuo para as organizações (como SAS). Em segundo lugar, o pipeline de inovação assume a forma de novas análises em desenvolvimento que são posteriormente adicionadas ao pipeline de produção.
- **Implementar comunicação e gerenciamento de processos:** notificações eficientes e automatizadas são essenciais em uma prática de DataOps. Quando alterações são feitas em qualquer código-fonte; ou quando um pipeline de dados é acionado, falha, concluído ou implantado, as partes interessadas certas podem ser notificadas imediatamente. Ferramentas para permitir a comunicação entre as partes interessadas também fazem parte do conjunto de ferramentas (pense no Slack ou no Trello).

No ambiente DataOps, o conjunto de ferramentas é muito importante para o sucesso, elas iram reduzir os contratempos até a interoperabilidade. apesar de integrar suas ferramentas em uma única base ou formar um conjunto de interoperáveis, essas tecnologias funcionam em todos os ambientes de dados atuais e previstos: local, nuvem, multi nuvem e híbrido. Na imagem abaixo (figura 6) mostra o fluxo do funcionamento das ferramentas.

Figura 6 - Fluxograma das ferramentas DataOps



Fonte: Autor

Os metadados inteligentes são essenciais no conjunto de ferramentas para o entendimento do que eles são, combinando ferramentas inteligentes com IA e ML abrangentes, os metadados inteligentes serão usados para melhorar a interferência. Ao criar metadados automaticamente na ingestão, detectá-los automaticamente no tempo de execução e rotular corretamente os objetos de dados, você pode reduzir significativamente o trabalho manual de sua equipe, tendo o desenvolvimento de pipelines de dados simplificando e acelerando análises eficazes.

Um exemplo muito bom é o impacto positivo da abordagem DataOps na detecção de fraudes, especialmente no setor bancário, as ferramentas na prática DataOps permitem que os bancos substituam a abordagem tradicional baseada em regras para detecção de fraudes para cartões de crédito de clientes, bem como aumentam consideravelmente a velocidade da análise de dados. Devido a esses insights avançados, os bancos podem evitar ter que ligar ou enviar mensagens de texto para seus clientes para verificar as transações.(ZAINO, 2019)

4.3 Vantagens

A implementação de uma prática DataOps pode trazer muitos benefícios para as corporações. Isso vai depender do processo, do projeto e das necessidades que cada empresa tem.

As empresas já envolvidas em DataOps concordam que esse paradigma terá um impacto positivo na organização. Embora as melhorias na agilidade e eficiência estejam intimamente relacionadas aos DataOps, os principais fatores, benefícios e prioridades estão, na verdade, relacionados à segurança e às mudanças transformacionais. O ambiente DataOps é mais avançado em termos de transição para a nuvem e execução de estratégias de transformação digital. Portanto, eles estão em uma posição melhor do que seus concorrentes.

Empresas que adotaram DataOps também obtêm proveito de muitas vantagens, portanto, além de mudanças no processo e na organização, também aposta em maior investimento em produtos e serviços. Dessa forma, os resultados da pesquisa confirmaram ainda mais nossa visão de que, embora DataOps ainda seja um termo quase desconhecido hoje, podemos esperar que nos próximos anos ele terá um impacto crescente em todo o mercado.

Vantagens gerais DataOps (MJV, 2020):

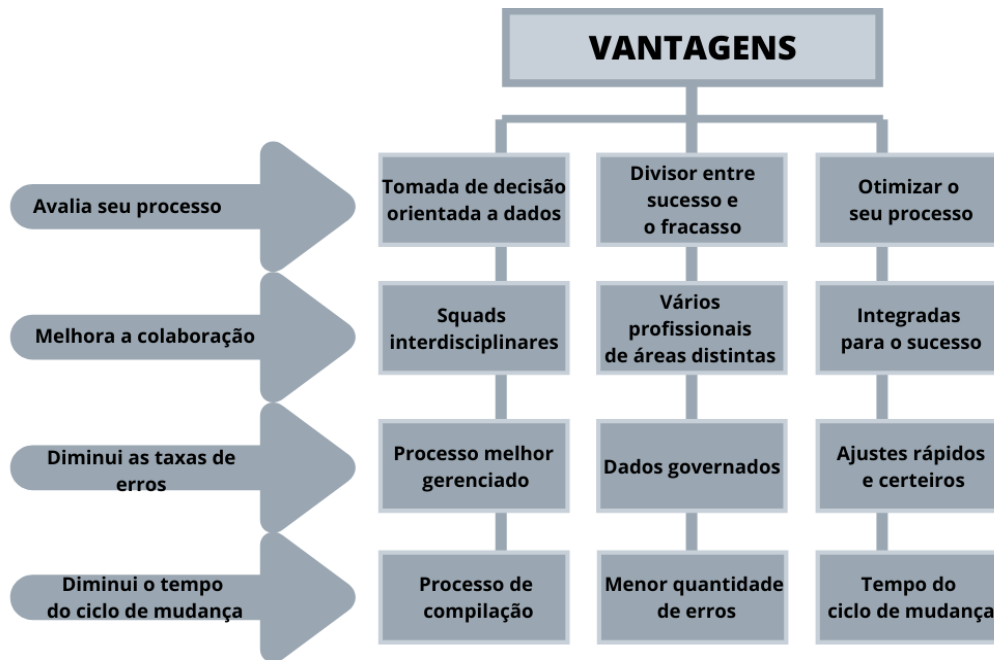
- **Avalia seu processo:** Já falamos aqui como o Data Science vem se tornando uma grande estrela nas corporações. A vasta maioria delas já percebe a importância e os benefícios da tomada de decisão orientada a dados. Inclusive, esse pode ser o divisor de águas entre o sucesso e o fracasso nesses tempos em que dados são inevitáveis. No entanto, quando falamos das atividades relacionadas aos processos de uma empresa, muitas vezes os processos são surpreendentemente não analíticos. Por isso que, dentro de uma estratégia baseada em DataOps, otimizar o seu processo ligado a dados (ou criá-lo, caso não exista) vai garantir uma entrega mais eficiente e vantagem competitiva.
- **Melhora a colaboração:** DataOps versa diretamente sobre squads interdisciplinares. Os squads funcionam como time interdisciplinares in loco ou em um esquema offshore/nearshore, basicamente você terá vários profissionais de áreas distintas trabalhando ao mesmo tempo em fases diferentes do seu projeto. O desafio reside no

fato de que suas atividades precisam estar sempre integradas para o sucesso do projeto. É essa troca rica de ponta a ponta, inclusive com o próprio cliente, que vai gerar valor para o negócio. A palavra de ordem aqui é colaboração.

- **Diminui as taxas de erros:** Com DataOps a taxa de erro diminui porque passa a existir um processo melhor gerenciado, mais automatizado, com dados governados, versionamento e melhoria contínua no que tange aos dados. Dessa forma, é possível identificar erros existentes e prever outros que possam surgir no futuro, possibilitando ajustes rápidos e certos, e elevando consideravelmente a qualidade.
- **Diminui o tempo do ciclo de mudança:** Sabemos que, em muitas companhias, o longo tempo de ciclo é um dos motivos para a análise falhar em melhorar a tomada de decisão orientada por dados. Isso ocorre porque novas análises dependem de processos obsoletos, manuais e, conseqüentemente, demorados. A criação de um processo de compilação sem intervenção manual – e com menor quantidade de erros – acelera exponencialmente o tempo do ciclo de mudança. Testes que validam e monitoram novas análises permitem implantar com confiança. E quando você tem certeza, pode implantar e integrar novas análises de forma muito mais rápida e eficiente.

Abaixo (Figura 7) o fluxo das vantagens nos processos DataOps:

Figura 7 - Fluxograma das vantagens DataOps



Fonte: autor

A beleza do DataOps reside no fato de que ele pode ser abraçado por toda a corporação. Desde o chairman até o colaborador. Não existe limitação para aplicar o DataOps e começar a ver melhoria na qualidade e na produtividade (IBIZ, 2020).

5 CONCLUSÃO

Este trabalho forneceu informações sobre DataOps, com o principal objetivo de mostrar os passos e etapas para implantação de um ambiente DataOps para incentivar a adoção desta prática em empresas.

Esse paradigma tem ganhado popularidade nas indústrias mundiais com sua capacidade de acelerar a produção de alta qualidade de insights de dados. Empresas como Google, Uber, Airplane utilizam DataOps como seu principal ativo para uma grande competição por inovação.

Porém, não tão conhecido como gostaríamos. No Brasil, devido à uma escassez de materiais de referência muito grande, IBM, iMasters e Data Science Academy são as mais confiáveis, no entanto seus centros de comando estão no exterior do país. Por ser um paradigma novo, obviamente teria a falta de materiais bibliográficos brasileiros, mas a falta de interesse de profissionais utilizando esta prática em empresas brasileiras demonstra desavanço por parte de empresas na competição de mercado.

As corporações sabem que a cada dia o volume de dados crescem e dados são de muita importância para o negócio, o que desafia o DataOps é fazer com que consumidores e operadores consigam trabalhar juntos, fornecendo dados de forma rápida, mas sem abrir mão da governança, segurança e qualidade dos dados.

DataOps objetiva a melhora de resultados, reunindo aqueles que precisam de dados com aqueles que o fornecem, eliminando o atrito ao longo do ciclo de vida dos dados permitindo o gerenciamento rápido, automatizado e seguro de dados.

Adotando este paradigma é preciso mudar práticas antiquadas de TI e inovar com novas metodologias e objetivos sem necessariamente precisar sair das ferramentas conhecidas, mas utilizadas de modo diferente. Este trabalho é um “ponta pé” inicial para mais materiais informativos, explicativos e explorativos, incentivando não só empresas como também profissionais da áreas de TI a se desafiarem com o que DataOps é capaz de fazer. Não importa de qual área da tecnologia da informação o profissional seja, contanto que a equipe tenha um pouco de tudo, com DataOps o sucesso é certo.

REFERÊNCIA BIBLIOGRÁFICA

ATWAL, Harvinder. **Practical DataOps: Delivering Agile Data Science at Scale**. Lean Thinking. Isleworth, Reino Unido: 2019 p.74

CISCO. **Cisco Annual Internet Report (2018–2023) White Paper**. Março de 2020 Disponível em:
<https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html>. Acesso em: 08 de outubro de 2020

COMPUTERWORLD. **Conheça a metodologia DataOps, que promete segurança e agilidade**. Fevereiro 2018. Disponível em:
<https://computerworld.com.br/plataformas/conheca-metodologia-dataops-que-promete-seguranca-e-agilidade/#:~:text=O%20DataOps%20adiciona%20fun%C3%A7%C3%B5es%20de,opera%C3%A7%C3%B5es%20e%20especialistas%20em%20dados>. Acesso em: 14 de outubro de 2020.

CYSEC BOT. **The DataOps Revolution: Delphix 2018 predictions**. Ano: 2018. Disponível em:
<https://www.maritimecybersecurity.center/the-dataops-revolution-delphix-2018-predictions/> Acessado em 25 de Outubro de 2020.

DATA KITCHEN. **The Seven Steps to Implement DataOps**. Call to action. 2017. p.6

DATA SCIENCE ACADEMY. **Engenheiro DataOps - Uma nova função no universo da ciência de dados**. Março de 2020. Disponível em :
<http://datascienceacademy.com.br/blog/engenheiro-dataops-uma-nova-funcao-no-universo-da-ciencia-de-dados/#:~:text=DataOps%20%C3%A9%20uma%20metodologia%20automatizada,ciclo%20da%20an%C3%A1lise%20de%20dados> Acesso em: 08 de outubro de 2020.

IBM. **IBM DataOps, Organize seus dados para que sejam confiáveis e prontos para os negócios, a fim de serem aplicados à sua jornada para a IA**. Ano: 2020 Disponível em:
<https://www.ibm.com/br-pt/analytics/dataops>. Acesso em: 03 de outubro de 2020.

IBIZ. **DATAOPS, um novo conceito que veio para ficar**. Agosto de 2020. Disponível em:
<http://www.ibiz.com.br/blogsol/dataops-um-novo-conceito-que-veio-para-ficar/> Acesso em: 27 de outubro de 2020.

GUPTA, Ritesh. IBM, **Components of the DataOps toolchain and best practices to make it successful**. Dezembro de 2019. Disponível em:
<https://www.ibm.com/blogs/journey-to-ai/2019/12/components-of-the-dataops-toolchain-and-best-practices-to-make-it-successful/> Acesso em: 26 de outubro de 2020.

LIEBMANN, Lenny. **3 reasons why DataOps is essential for big data success**. Junho de 2014. Disponível em:
<https://www.ibmbigdatahub.com/blog/3-reasons-why-dataops-essential-big-data-success> Acesso em: 03 de outubro de 2020

LOPES, Alex. **O que é DataOps: Organizando o futuro do Data Science para negócios.** Fevereiro de 2019. Disponível em: <https://medium.com/data-hackers/o-que-%C3%A9-dataops-organizando-o-futuro-do-da-ta-science-para-neg%C3%B3cios-dc46af338e12> Acesso em: 08 de outubro de 2020.

MANIFESTO. **O manifesto DataOps.** Disponível em: <https://www.dataopsmanifesto.org/> Acesso em: 03 de outubro de 2020.

MATSUDA, Gabriel. **O que você precisa saber sobre DataOps e por que você deveria se importar?** Janeiro de 2020 Disponível em: <https://www.baguete.com.br/noticias/10/01/2020/o-que-voce-precisa-saber-sobre-dataops-e-por-que-voce-deveria-se-importar> Acesso em: 03 de outubro de 2020.

MEDIUM. **DataOps is NOT Just DevOps for Data .** Novembro de 2018. Disponível em: <https://medium.com/data-ops/dataops-is-not-just-devops-for-data-6e03083157b7> Acesso em: 19 de outubro de 2020

MJV. **DataOps: O futuro chegou para ficar.** 4 Vantagens do DataOps. 2020. p.16.

VECTOR ITC. **DataOps ou como lidar com dados de forma eficaz e eficiente.** Agosto de 2017. Disponível em: <https://www.vectoritcgroup.com/pt-pt/tech-magazine-pt/data-ecosystem-pt-pt/dataops-ou-como-manipular-dados-de-forma-eficaz-e-eficiente/> Acesso em: 26 de Outubro de 2020.

ZAINO, Jennifer. HPE. **How to get started with DataOps.** Junho de 2019. Disponível em: <https://www.hpe.com/us/en/insights/articles/how-to-get-started-with-dataops-1906.html> Acesso em: 27 de outubro de 2020.