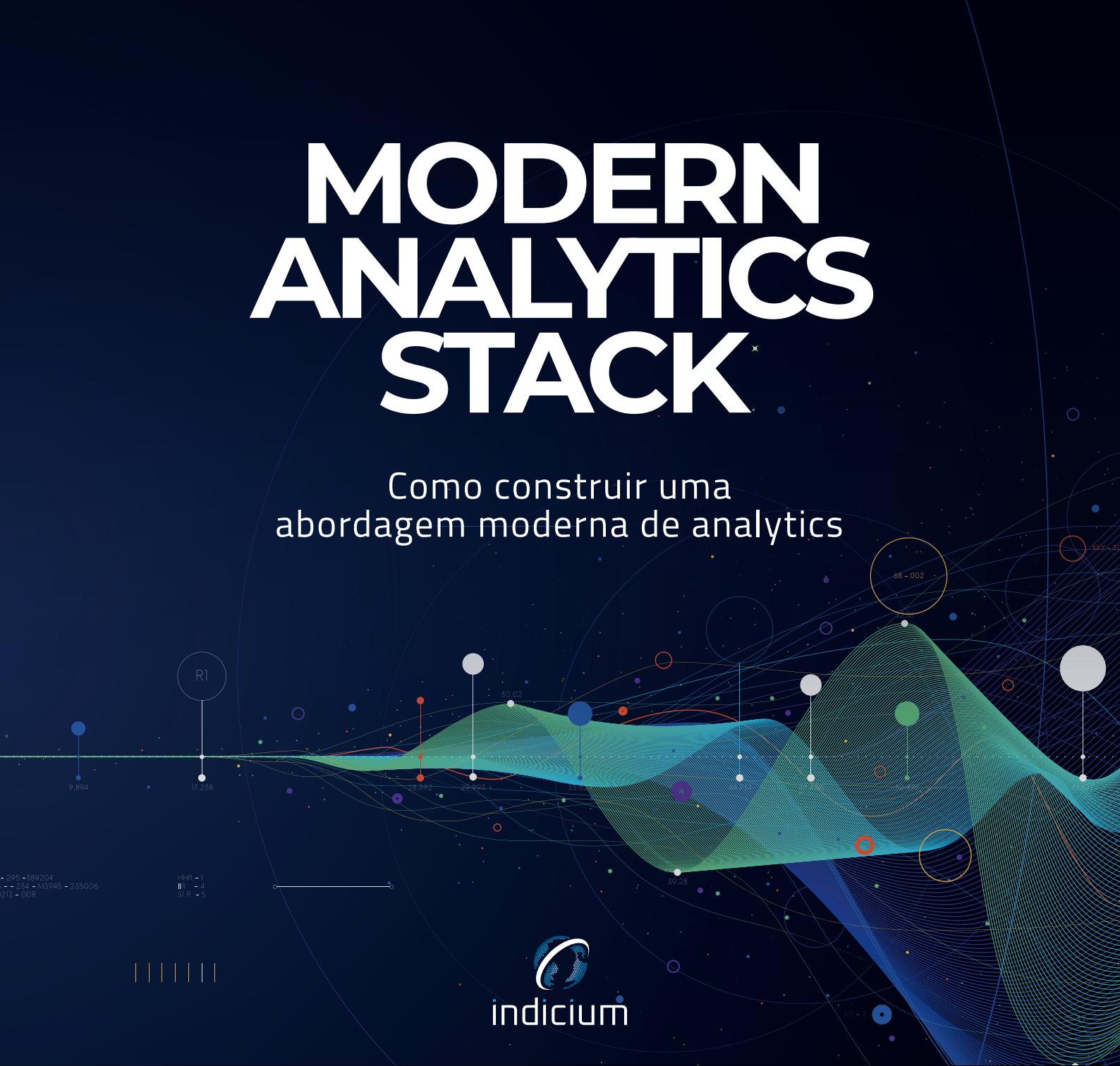


MODERN ANALYTICS STACK

Como construir uma
abordagem moderna de analytics



295 - 289204
134 - MS945 - 285006
121 - DDR

HHR - 1
III - 4
SIR - 5



SUMÁRIO

1. Introdução	00
2. Os desafios das operações modernas de dados	00
3. Modern analytics stack: o que é?	00
4. Os princípios da abordagem moderna de analytics	00
5. Construindo uma abordagem moderna de analytics	00
5.1. Data stack: os 5 estágios da pilha de dados	00
· Coleta, limpeza e integração de dados	00
· Armazenamento e gerenciamento de dados	00
· Transformação de dados	00
· Business intelligence e data analytics	00
· Advanced analytics	00

6. Data Stack: as principais ferramentas da abordagem moderna de analytics

00

6.1. Ferramentas de coleta, deployment e transformação de dados

00

- Coleta e integração de dados 00
- Deployment 00
- Transformação de dados 00

6.2. Ferramentas de armazenamento e gerenciamento de dados

00

- Data warehouses 00
- Data lakes 00
- Soluções de computação na nuvem 00

6.3. Ferramentas de data analytics

00

- Business intelligence 00
- Advanced analytics 00

Conclusão

00

INTRODUÇÃO

Graças às novas tecnologias da era do big data e da computação na nuvem, houve uma redução drástica nos custos de processamento, possibilitando a exponencialização de ferramentas complexas, capazes de processar volumes abundantes de dados.

Com isso, algoritmos, modelos estatísticos e preditivos tornaram-se muito mais acessíveis e, desde então, a inteligência de dados virou uma grande **oportunidade de negócio** e, sobretudo, uma fonte de diferenciação competitiva nas organizações modernas.

Em razão dessas mudanças, há novas abordagens flexíveis surgindo para auxiliar organizações a escalar suas operações de dados que você precisa conhecer, como a **abordagem moderna de analytics**.

Neste e-book, você vai aprender tudo o que precisa saber sobre a abordagem moderna de analytics, incluindo o que é, quais os seus princípios e as suas características. Para completar, terá a oportunidade de descobrir como construir o seu data stack e implementá-lo com sucesso, utilizando ferramentas de última geração para revolucionar o gerenciamento de dados na sua empresa.



OS DESAFIOS DAS OPERAÇÕES DE DADOS MODERNAS

Nas próximas décadas, o volume, a velocidade e a variedade de dados tomarão proporções astronômicas. De acordo com o IDC, a esfera global de dados deverá crescer 500% até 2025. Como consequência, dados nunca foram tão fundamentais no meio empresarial.

Hoje, dados sobre o comportamento do cliente, estoques de produto, informações de mercado e muito mais, podem ser rastreados e analisados em tempo real para fornecer percepções críticas aos negócios. Dessa forma, as empresas têm o poder de se tornarem incrivelmente ágeis e identificarem sinais em seus dados em tempo real, podendo otimizar resultados e tomarem decisões eficientes para responderem rapidamente aos dilemas dos negócios.

Porém, à medida que mais dados são gerados, novos desafios quanto à organização, à integração e ao gerenciamento dessas informações surgem. Dentre eles:

- 
- a demanda por sistemas de armazenamento e computação na nuvem mais eficientes;
 - a necessidade de integração de arquiteturas de dados;
 - a demanda pela automatização de tarefas rotineiras com auxílio da inteligência artificial etc.

Percebe-se, portanto, que as soluções de gerenciamento de dados estão **sob pressão** para serem cada vez mais rápidas, eficientes e capazes de comportar grandes volumes de informações. Mas, além disso, também devem ser flexíveis para incorporar diversas versões da verdade e precisam, sobretudo, valorizar e facilitar a experiência do usuário final para serem verdadeiramente bem-sucedidas.

Na prática, ao passo que as empresas aumentam o seu tamanho e, naturalmente, a complexidade das suas operações de dados, as metodologias e os sistemas, que uma vez funcionaram em pequena escala, tornam-se obsoletos e passam a causar atritos prejudiciais aos negócios, devendo, por fim, ser substituídos.

Nesse contexto, surge a seguinte indagação: como uma empresa pode preparar a sua operação e infraestrutura de dados para lidar com tantos desafios?

A resposta para reduzir tais impactos na prática empresarial é a **implementação de abordagens modernas de gerenciamento de dados** que ajudam a democratizar a coleta, integração e o gerenciamento de dados para todas as partes interessadas relevantes.

Em resumo, a solução para as empresas se prepararem para

A dark blue background featuring a complex network of glowing pink and purple lines and dots, resembling a star field or a neural network.

operações de dados cada vez mais complexas é a construção de metodologias de gerenciamento de dados dinâmicas, mutáveis e flexíveis, como a **abordagem moderna de analytics**.

Continue a leitura e saiba tudo sobre o assunto.



MODERN ANALYTICS STACK: O QUE É?

A abordagem moderna de analytics (MAS) tem vários nomes diferentes:

- [modern analytics stack](#) (MAS)
- [modern data stack](#) (MDS)
- [analytics stack](#)
- data stack

Apesar das distinções de nomenclatura, essa metodologia surgiu para resolver um problema central do mercado: auxiliar empresas a desenvolverem a sua **capacidade analítica** frente a desafios, como o avanço rápido de novas tecnologias, o aumento do volume de dados e a complexidade crescente dos negócios.

Mas, afinal, como a modern analytics stack se propõe a solucionar isso?

Para suportar o crescimento escalável das operações modernas de dados, essa abordagem engloba a combinação de [boas práticas de data science, ferramentas analíticas específicas e tecnologias inovadoras para a criação de infraestruturas de dados flexíveis, adaptáveis e acessíveis](#).



Por isso, muito mais que uma tecnologia isolada, a abordagem moderna de analytics reúne todos os elementos necessários para solucionar os inúmeros desafios de data science e analytics das empresas modernas.

Na prática, uma empresa que possui um data stack estruturado utiliza ferramentas e tecnologias independentes, mas complementares, que executam funções específicas ao longo do seu ciclo de dados.

Por exemplo, considere uma empresa que aumentou drasticamente o seu estoque e a sua base de clientes e, por isso, precisa de uma solução de armazenamento de dados mais robusta.

Nesse caso, graças ao data stack, ela poderá adequar a sua solução atual ou, ainda, substituí-la por uma nova que atenda às novas demandas do negócio. E o melhor: tudo isso **sem precisar reformular toda a sua infraestrutura de dados por completo**.

Com a MAS, assim como em um jogo de Lego, as organizações têm a flexibilidade e a autonomia para substituírem peças e fazerem ajustes pontuais, sem necessariamente destruírem toda a sua operação de dados.



OS PRINCÍPIOS DA ABORDAGEM MODERNA DE ANALYTICS

Para construir uma abordagem de analytics moderna e eficiente, é preciso conhecer a fundo suas principais características e seus princípios norteadores. Pensando nisso, elencamos abaixo os 7 princípios de um dos conceitos mais modernos de analytics. Confira!

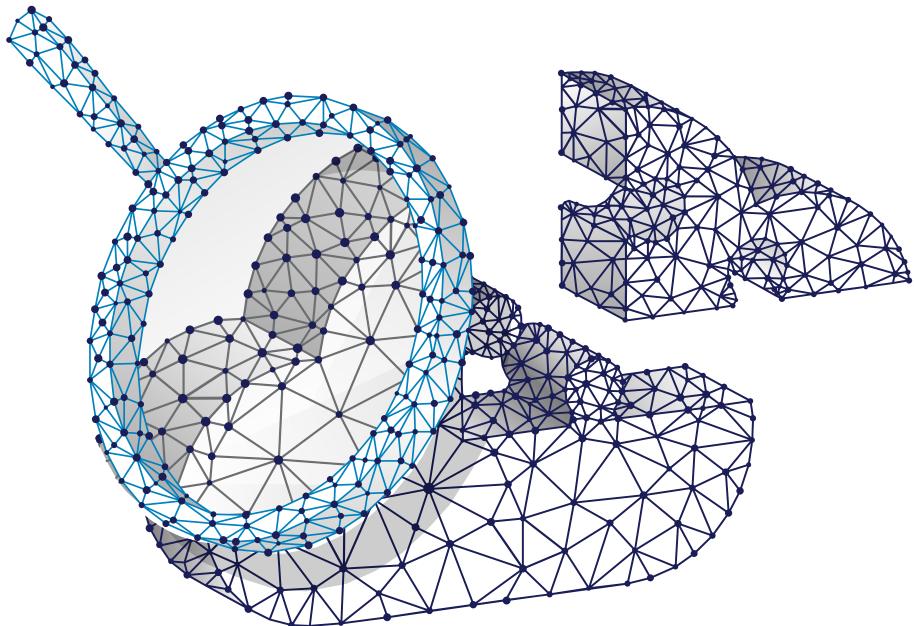
- 1. Cloud-based:** armazenamento de dados totalmente **cloud-based**, centralizado na nuvem, tecnologia altamente escalável e flexível que permite o processamento de uma quantidade virtualmente infinita de dados em um ambiente online e seguro, reduzindo custos com infraestrutura, instalação e manutenção.
- 2. Modularidade:** separação das regras de negócio das etapas do processo de **ELT**, particularmente nas fases de **extração (extraction)** e **carregamento (load)**, permitindo o uso de **third-party tools**, ferramenta de terceiros para dar seguimento ao processo de integração de dados com baixo investimento.
- 3. Simplicidade:** transformação de dados orientada pela simplicidade de uma ou poucas linguagens de programação de amplo conhecimento, feita de maneira centralizada, trazendo benefícios, como a democratização da informação e a redução de custos com treinamento e manutenção.

- 
- 4. Governança:** informações centralizadas e facilmente acessíveis em um só local, simplificando a documentação e governança dos dados, permitindo a criação de lógicas de permissionamento e o gerenciamento de dados sensíveis de forma integrada.
 - 5. Versionamento:** utilização de boas práticas de versionamento permitindo o trabalho colaborativo em projetos de dados, sem geração de conflitos graças às ferramentas modernas de ELT utilizadas nessa abordagem.
 - 6. Separação de ambientes:** criação de ambientes distintos para separação de dados brutos, dados em transformação e dados finais, facilitando o acesso a diferentes ambientes de desenvolvimento, o trabalho colaborativo e a redução de erros de produção
 - 7. Testes:** boas práticas de testes nos projeto de dados, como ocorre em projetos modernos de software, garantem a consistência e a confiabilidade dos resultados.

Agora que você conhece as características, vamos falar sobre como construir a abordagem moderna de analytics na sua empresa?

CONSTRUINDO UMA ABORDAGEM MODERNA DE ANALYTICS

Hoje, graças às novas tecnologias e ferramentas acessíveis, empresas de todos os portes podem adotar a abordagem moderna de analytics. No entanto, para obter sucesso na implementação dessas práticas, é necessário entender como todas as suas peças se encaixam, desde as etapas do data stack até as tecnologias e ferramentas recomendadas para a sua execução.





Data stack: os 5 estágios da pilha de dados

Uma das características mais marcantes da abordagem moderna de analytics é a **junção de diversas de ferramentas e tecnologias** em um data stack.

Mas, afinal, o que é um data stack?

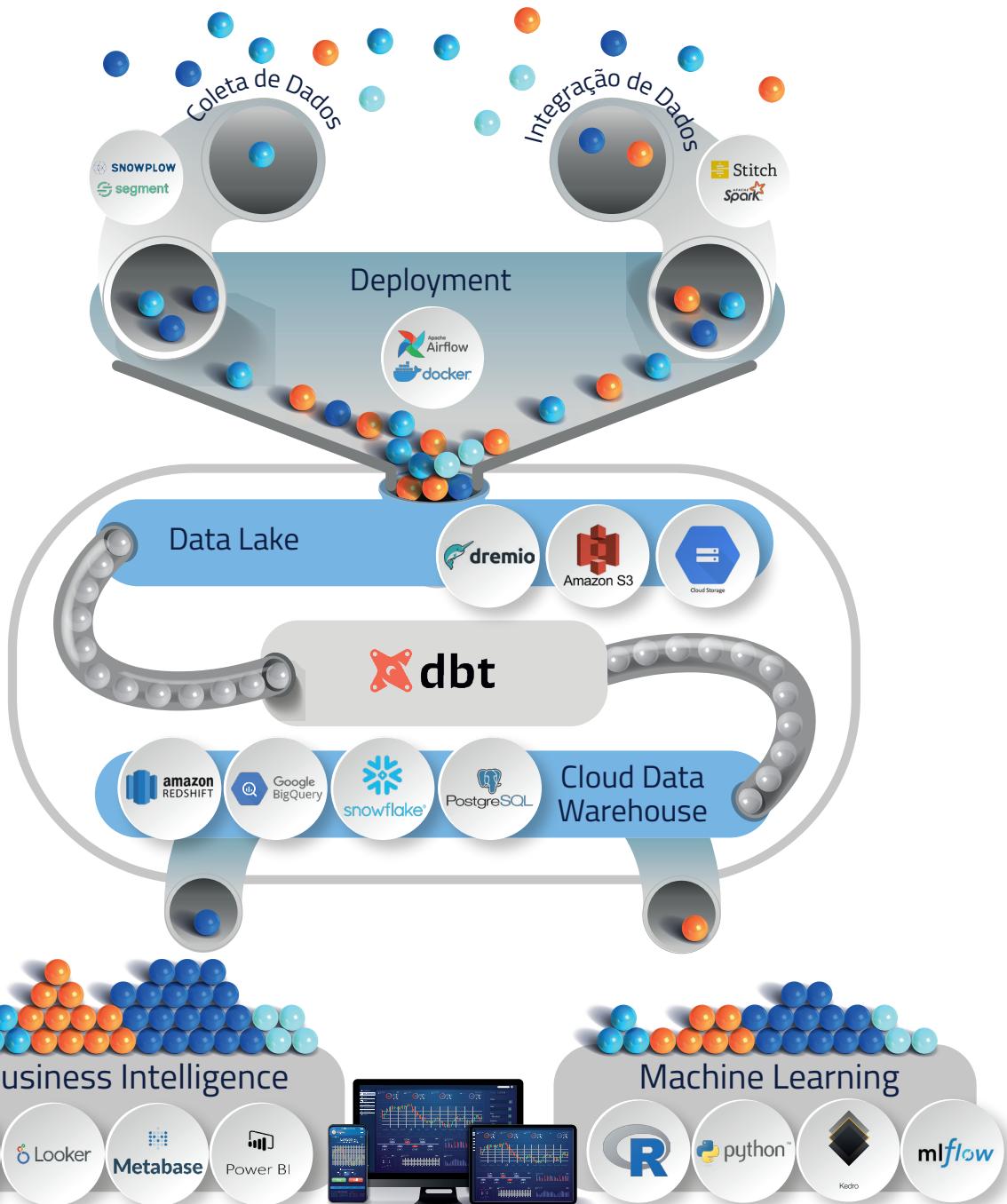
A pilha de dados, ou data stack, é a coleção de processos, ferramentas, aplicações e tecnologias responsável pela automatização do gerenciamento de dados no negócio em todas as etapas do data pipeline.

O primeiro passo na implementação da abordagem moderna de analytics, portanto, é a estruturação e, posteriormente, configuração do “data stack”. Com isso, é possível responder às demandas das operações de dados atuais, de forma altamente eficiente, ao longo do fluxo de dados.

Um data stack eficiente deve seguir etapas e atribuições específicas ao longo do data pipeline. Atualmente, as principais tarefas e ferramentas do MAS podem ser resumidas no diagrama a seguir.



5 estágios





ESTÁGIO 01

Coleta, integração e limpeza de dados: empresas têm seu próprio processo para coleta de dados relevantes. Nessa etapa do data stack, é possível coletar e integrar dados de múltiplas fontes como CRMs, planilhas de Excel, mídias sociais etc, centralizando-os em um data warehouse de forma eficiente. É, portanto, nesse momento, com auxílio das ferramentas corretas, que realizam-se os ajustes necessários para que os dados estejam preparados para a etapa seguinte do data stack.



ESTÁGIO 02

Armazenamento e gerenciamento de dados: agora, os dados devem ser preparados e armazenados em data lakes e data warehouses, estruturas escaláveis e seguras que possibilitam análises e o gerenciamento de informações em larga escala. Essas ferramentas são componentes fundamentais do data stack.



ESTÁGIO 03

Transformação de dados: considerando o volume massivo de dados para processamento, ao invés do tradicional ETL (extract-transform-load), no data stack, utiliza-se o fluxo de ELT, uma abordagem mais rápida e flexível para a transformação de dados na qual esse processo ocorre logo após a coleta e integração das informações em um repositório de dados centralizado, e não antes, como ocorre nos casos do ETL. Com isso, é possível transformar dados brutos em dados modelados dentro de um data warehouse.



ESTÁGIO 04

Business intelligence e data analytics: inteligência analítica é a prioridade final do modern analytics stack. Assim, depois da configuração das etapas anteriores, as informações e os insights que

agregam valor à tomada de decisão empresarial finalmente tornam-se acessíveis aos usuários finais em tempo real. Como resultado, gestores e departamentos de negócio podem visualizar dados, identificar tendências, otimizar processos e agir rapidamente com auxílio de ferramentas de business intelligence, dashboards interativos e relatórios inteligentes, conectados a um data warehouse.

Advanced analytics: na última e mais avançada etapa da MAS, é possível aplicar e desenvolver técnicas avançadas de machine learning, inteligência artificial e modelagem preditiva altamente complexas, como modelos de recomendação e modelagem prescritiva, dentro da infraestrutura moderna de dados configurada nas etapas anteriores.

Na prática, qualquer data stack construído com base nas 6 operações básicas descritas acima preenche os requisitos necessários para suportar o crescimento escalável das operações de dados modernas. Dessa forma, ainda que a arquitetura de um pipeline de dados varie de acordo com as empresas, todas devem ter esses processos incorporados.

Todos esses conceitos abordados nesta seção serão extremamente úteis à medida que avançamos na discussão sobre as ferramentas e tecnologias recomendadas para a implementação de um modern data stack no seu negócio.



DATA STACK: AS PRINCIPAIS FERRAMENTAS DA ABORDAGEM MODERNA DE ANALYTICS

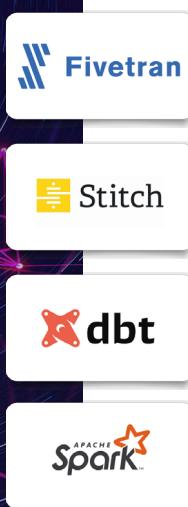
Uma das características mais marcantes da abordagem moderna de analytics é a junção de diversas **ferramentas e tecnologias** para responder às demandas das operações de dados atuais. Por isso, além das etapas do fluxo de dados, as ferramentas usadas em cada estágio do data stack são elementos essenciais que determinam o sucesso da implementação da modern analytics stack.

Uma operação de dados moderna requer, portanto, a combinação de diversos serviços e ferramentas em um data stack, conforme imagem a seguir.



Confira na sequência a análise das principais ferramentas disponíveis no mercado para a construção de uma abordagem moderna de analytics escalável e flexível na sua empresa.

FERRAMENTAS DE COLETA, DEPLOYMENT E TRANSFORMAÇÃO DE DADOS



Coleta e integração de dados: ferramentas como o [Fivetran](#) e [Stitchdata](#) são as líderes na integração de dados na nuvem. Elas permitem mover dados de centenas de fontes como ERPs, CRMs, bancos de dados, REST APIs etc. diretamente para um data warehouse (na nuvem ou on-premises). Além disso, podem ser combinadas. Com isso, não há necessidade de grandes investimentos em licenças de software ou horas de implementação.

Deployment: ferramentas como [Docker](#) e o [Kubernetes](#) são muito utilizadas para realizar o deployment em conjunto com orquestradores, como o [Airflow](#) e [Prefect](#). O diferencial dessas tecnologias é fazer com que todas as “peças do Lego” conversem de forma harmônica entre si, garantindo que os dados fluam em sintonia ao longo da estrutura de dados.

Transformação de dados: as três principais ferramentas do modern analytics stack usadas para a transformação de dados são o [DBT](#) (data build tool), o [Dataform](#) e o Spark. Juntas, essas tecnologias permitem a execução do processo de ELT para a transformação dos dados brutos em dados modelados dentro de um data warehouse, etapa fundamental da abordagem moderna de analytics.



FERRAMENTAS DE ARMAZENAMENTO E GERENCIAMENTO DE DADOS

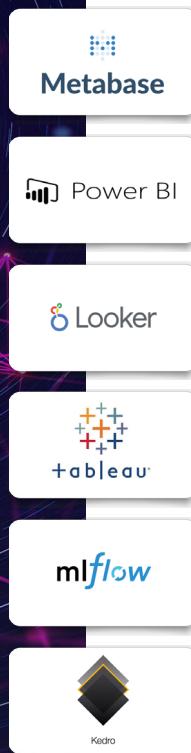
Hoje, existem duas opções práticas e eficientes quanto ao armazenamento de dados tanto on-premises quanto na nuvem: [data warehouses](#) e [data lakes](#). Ambas são alternativas viáveis, no entanto, devem ser avaliadas caso a caso, pois apresentam diferenças técnicas e conceituais em termos de arquitetura e finalidade.

Data warehouses: a arquitetura escalável de data warehouses na nuvem, como o Amazon Redshift, Snowflake, Google Big Query e Post Gre SQL, permite o armazenamento e a consulta de volumes enormes de dados de forma rápida. Por isso, essas são ferramentas essenciais na construção de uma abordagem moderna de analytics.

Data lakes: armazenam todos os tipos de dados - estruturados, não estruturados e híbridos - em um só lugar. Para essa finalidade, indicam-se as seguintes ferramentas: Dremio e Amazon S3.

Computação na nuvem: atualmente, os principais provedores de computação na nuvem são a AWS, Google Cloud e Microsoft Azure. É possível escolher quaisquer dessas opções de acordo com a necessidade da sua empresa para o armazenamento e gerenciamento dos seus dados.

FERRAMENTAS DE DATA ANALYTICS



Data analytics envolve diversas atividades quanto à análise de dados, que variam em termos de graus de complexidade e ferramentas. Para facilitar a sua compreensão, é possível dividi-las duas categorias: **ferramentas de business intelligence** e **advanced analytics**, ou análises avançadas.

Business intelligence: com uma infraestrutura de dados moderna estabelecida, é possível utilizar diferentes ferramentas de business intelligence para visualizar, analisar e gerar insights a partir dos dados. Para tanto, há diversas alternativas open-source robustas, como o Metabase, bem como plataformas SaaS, como o Microsoft PowerBI, Looker, Tableau entre outros.

Advanced analytics: machine learning, inteligência artificial e técnicas de modelagem são técnicas aplicadas em advanced analytics para análises mais complexas dentro do data stack. Para isso, além das diversas bibliotecas nas linguagens R e Python, ferramentas como MlFlow e Kedro auxiliam na execução de modelos preditivos, prescritivos e, também, otimizam o processo de desenvolvimento, reduzindo o tempo entre a modelagem e a utilização do usuário final.

CONCLUSÃO

Existem muitas decisões envolvidas no desenvolvimento e aprimoramento da operação de dados em âmbito empresarial. Informações atualizadas podem ser difíceis de encontrar e, muitas vezes, ainda mais difíceis de entender.

A abordagem moderna de analytics surge para solucionar esse problema, criando uma metodologia integrada de inteligência de negócios que coleta, combina, analisa e entrega o valor dos dados para empresas.

Felizmente, os componentes do data stack estão muito mais baratos, simples de configurar e manusear. Assim, empresas de todos os portes podem utilizá-lo para obter vantagem competitiva e desenvolver maturidade analítica.

Se você quer aprender mais sobre como empresas de todos os tamanhos estão modernizando seus programas de gerenciamento de dados, entre em contato conosco. Podemos ajudar você a entender suas necessidades atuais de gerenciamento de dados, avaliar as suas opções e indicar como dar os próximos passos.



indicium

Somos uma empresa de ciência de dados e analytics, [**Líder na América Latina**](#), formada por profissionais com expertise multidisciplinar e experiência de mercado.

Promovemos a **redução de custos e o aumento de receita com técnicas estatísticas, big data, machine learning, inteligência artificial e os principais softwares para garantir a melhor adequação de empresas na Jornada Data Driven.**

Entregamos informações claras e objetivas para a formulação de estratégias, a configuração de infraestrutura, a elaboração de indicadores e o desenvolvimento de modelos preditivos, possibilitando o **fácil acesso à informação para tomada de decisões em tempo real.**

www.indicium.tech



indiciumtech



indiciumtech



indicium_tech



indiciumtech