

[E-BOOK]

MODERN DATA STACK

Como construir uma plataforma moderna de dados.



SUMÁRIO

1. Introdução	3
2. Desafios das operações modernas de dados	4
3. O que é Modern Data Stack?	7
4. Os princípios da abordagem moderna de dados	9
5. Construindo uma abordagem moderna de dados	11
6. Data stack: principais ferramentas da abordagem moderna de dados	16
6.1 Ferramentas de coleta, deployment e transformação de dados	18
6.2 Ferramentas de armazenamento e gerenciamento de dados	19
6.3 Ferramentas de data analytics	20
7. Conclusão	21
Conheça a Indicium	22

1 | Introdução

Modern Data Stack é a abordagem mais moderna para se trabalhar com dados, mas as coisas não evoluíram do dia para a noite.

Com o surgimento das novas tecnologias de **big data** e **computação em nuvem**, os custos de processamento de dados sofreram uma redução drástica.

Isso possibilitou o desenvolvimento de **ferramentas de dados** cada vez mais complexas, capazes de processar volumes abundantes de dados.

Algoritmos, modelos estatísticos e preditivos foram ficando mais acessíveis e, agora, a **inteligência de dados** é uma grande **oportunidade de negócio**.

Sobretudo, é uma **fonte de diferenciação competitiva** para as empresas.

E, com essas inovações, surgiram **novas abordagens e metodologias mais flexíveis** para se trabalhar com os dados e auxiliar organizações na escalada de suas operações.

Este e-book é para você conhecer tudo sobre a **abordagem moderna de dados** que adotamos aqui na **Indicium** para todos os nossos clientes e parceiros. Você vai entender o que é **Modern Data Stack (MDS)**, quais os princípios e as características.

Mais que isso, você vai aprender a como construir o sua **data stack** e implementá-la com sucesso utilizando ferramentas de última geração para revolucionar o gerenciamento de dados na sua empresa.

Aproveite!



2 | Desafios das operações modernas de dados

Nas próximas décadas, **o volume, a velocidade e a variedade de dados** tomarão proporções astronômicas.

De acordo com o International Data Corporation (IDC), a esfera global de dados deverá crescer 500% até 2025. Como consequência, os dados nunca foram tão fundamentais no meio empresarial como são hoje.

Dados sobre comportamentos de clientes, estoques, produtos, mercado e muito mais, podem ser rastreados e analisados em tempo real para fornecer percepções críticas aos negócios.

E é assim que as empresas estão se tornando incrivelmente ágeis.

Identificando sinais em seus dados em tempo real, os resultados são otimizados e as decisões tomadas são muito mais eficientes para responderem rapidamente os dilemas dos negócios.

Porém, nem tudo são flores...

Quanto mais dados são gerados, novos desafios surgem sobre a organização, integração e o gerenciamento deles, como por exemplo a necessidade de::



sistemas de armazenamento e computação em nuvem mais eficientes



integração de arquiteturas de dados



automatização de tarefas rotineiras com auxílio da inteligência artificial.



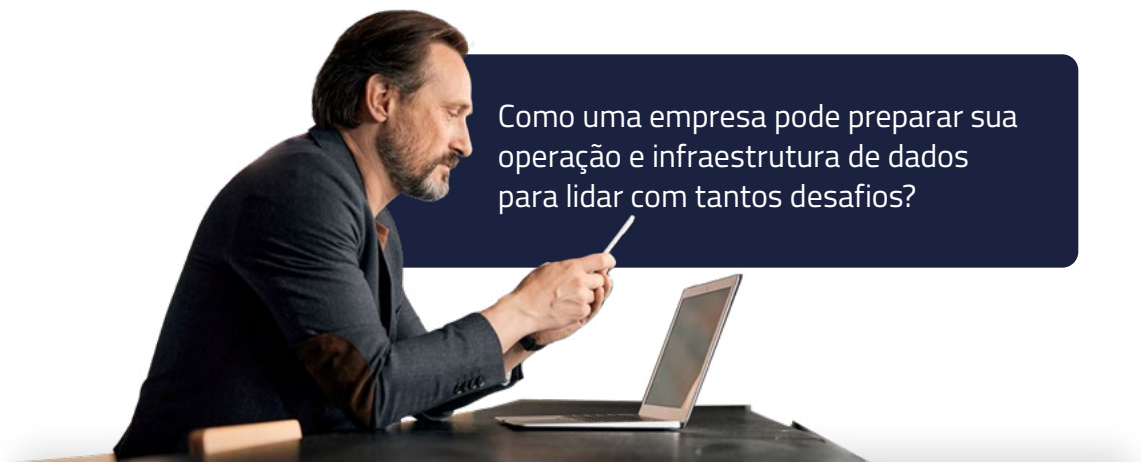
Além disso, as soluções de gerenciamento de dados estão sob pressão para serem:

- cada vez **mais rápidas**, eficientes e capazes de comportar grandes volumes de informações;
- cada vez **mais flexíveis** para incorporar diversas versões da verdade;
- cada vez **mais acessíveis** para valorizar e facilitar a experiência do usuário final sendo 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 uma dúvida:



Como uma empresa pode preparar sua operação e infraestrutura de dados para lidar com tantos desafios?

A resposta para isso é: implementar uma abordagem moderna de gerenciamento de dados, ou seja, a **Modern Data Stack (MDS)**, uma solução inovadora que democratiza a coleta, a integração e o gerenciamento de dados para todas as partes interessadas e relevantes.

Resumindo...

Uma empresa precisa se preparar para operações de dados cada vez mais complexas **construindo plataformas de gerenciamento de dados** dinâmicas, mutáveis e flexíveis, como a **abordagem moderna de dados que usamos na Indicium**.

Você vai conhecê-la bem a seguir!



3 | O que é Modern Data Stack?

A **abordagem moderna** de dados que adotamos na **Indicium** é conhecida por vários nomes:

- **Modern Data Stack (MDS)**
- Modern Analytics Stack
- Data Stack
- Analytics Stack

Mas, apesar das nomenclaturas diferentes, estamos falando de uma mesma metodologia que surgiu para resolver um problema central do mercado: **desenvolver a capacidade analítica das empresas** frente aos desafios do dia a dia.

Que tipos de desafios?

O avanço rápido de novas tecnologias, o aumento do volume de dados e a complexidade crescente dos negócios, por exemplo.

3.1 | Como a Modern Data Stack soluciona isso?

Para **suportar o crescimento escalável das operações modernas de dados**, com essa abordagem nós criamos **infraestruturas de dados flexíveis, adaptáveis e acessíveis** com a combinação de:

- boas práticas de data science;
- ferramentas analíticas específicas; e
- tecnologias inovadoras.



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

NA PRÁTICA...

Uma empresa com uma Modern Data Stack estruturada 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 à **data stack**, ela poderá adequar a sua solução atual ou, ainda, substituí-la por uma que atenda às novas demandas dos seus negócios. E, o melhor: sem precisar reformular toda a sua infraestrutura de dados por completo.

Resumindo...

Com a MDS, 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 estrutura de dados.



4 | Os princípios da abordagem moderna de dados

Para construir uma **abordagem de dados moderna e eficiente**, é preciso conhecer a fundo suas principais características e seus princípios norteadores.

Pensando nisso, elencamos **sete princípios** de um dos conceitos mais modernos da área de dados. Confira!

1º | Cloud based

Armazenamento de dados totalmente centralizado em nuvem, com tecnologia escalável e flexível que permite o processamento de uma quantidade infinita de dados em um ambiente online e seguro, reduzindo custos com infraestrutura, instalação e manutenção.

2º | Modularidade

Separação das regras de negócio das etapas do processo de **ELT**, particularmente nas fases de extração e carregamento, permitindo o uso de ferramenta de terceiros (third-party tools) para dar seguimento ao processo de integração de dados com baixo investimento.

3º | Simplicidade

Transformação de dados orientada por 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.

4º | Governança

Informações centralizadas e facilmente acessíveis, simplificando a documentação e governança dos dados, permitindo a criação de lógicas de permissionamento e o gerenciamento de dados sensíveis de forma integrada.



5º | 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.

6º | 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.

7º | Testes

Execução de boas práticas de testes nos projeto de dados, como ocorre em projetos modernos de software, garantindo a consistência e a confiabilidade dos resultados.

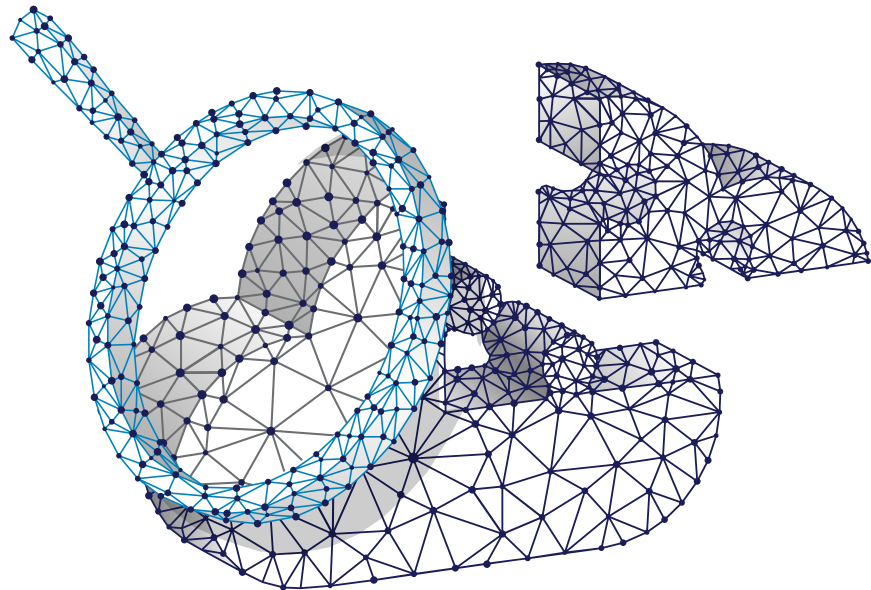
Agora que você conhece esses princípios fundamentais da **abordagem moderna de dados**, vamos falar sobre como você pode construir isso na sua empresa.



5 | Construindo uma abordagem moderna de dados

Graças às novas tecnologias e ferramentas acessíveis, empresas de todos os portes podem adotar uma **abordagem moderna de dados** hoje em dia.

Mas é necessário entender como todas as suas peças se encaixam para ter sucesso na implementação dessas práticas, desde as **etapas da data stack** até a definição das tecnologias e ferramentas para sua execução.





5.1 | Data stack: cinco estágios da pilha de dados

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

• Você sabe o que é data stack?

A **pilha de dados** ou **data stack** é uma **coleção de processos, ferramentas, aplicações e tecnologias** responsável pela automatização do **gerenciamento de dados** nos negócios em todo o **pipeline de dados**.

• Você sabe o que é pipeline de dados?

Traduzindo **pipeline**, temos gasoduto ou tubulação em português, e na nossa **área de dados, pipeline são todas as etapas do processamento de dados**, desde o sistema de entrada até o destino final da informação, ou seja, é o complexo todo desse processo.



NA PRÁTICA...

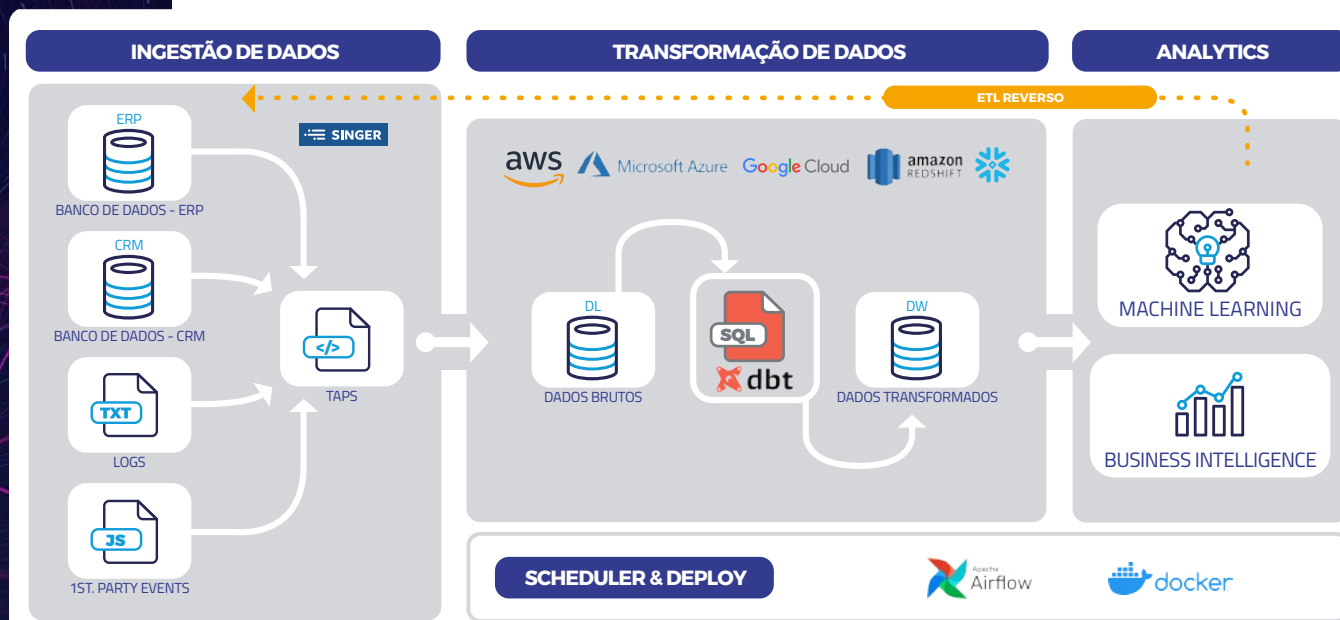
O primeiro passo na implementação da MDS é a estruturação.

Depois, vem a configuração da data stack.

Com isso, é possível responder demandas das operações de dados de forma altamente eficiente.



Para sua empresa ter uma **data stack** eficiente, há cinco estágios específicos ao longo do **pipeline de dados**. E as principais tarefas e ferramentas da **Modern Data Stack** estão resumidas no diagrama a seguir em suas respectivas etapas.





1º ESTÁGIO

Coleta, integração e limpeza de dados

Empresas têm seu próprio processo para coleta de dados relevantes. Nesta etapa da **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 nesse momento, com auxílio das ferramentas corretas, que são feitos os ajustes necessários para que os dados estejam preparados para a etapa seguinte.

2º ESTÁGIO

Armazenamento e gerenciamento de dados

Agora, os dados são preparados e armazenados em **data lakes** e **data warehouses**, estruturas escaláveis e seguras que possibilitam análises e o gerenciamento de informações em larga escala, e que são fundamentais na **data stack**.

3º ESTÁGIO

Transformação de dados

Aqui, por conta do volume massivo de dados para processamento, ao invés do tradicional **ETL** (extrair, transformar e carregar), na **Modern Data Stack**, utilizamos o fluxo de **ELT**, uma abordagem mais rápida e flexível para a **transformação de dados** que ocorre logo após a coleta e integração das informações em um repositório centralizado, e não mais antes, como no ETL. Com isso, é possível transformar dados brutos em dados modelados dentro de um **data warehouse**.

4º ESTÁGIO

Business intelligence e data analytics

A **inteligência analítica** é a prioridade final da **Modern Data Stack**. Assim, depois da configuração das etapas anteriores, as informações e os insights que agregam valor à tomada de decisão empresarial finalmente se tornam acessíveis aos usuários finais em tempo real. Como resultado, gestores e departamentos de negócios podem visualizar dados, identificar tendências, otimizar processos e agir rapidamente com auxílio de ferramentas de **business intelligence**, dashboards interativos e relatórios automatizados, conectados a um **data warehouse**.



5º ESTÁGIO

Advanced analytics

No último e mais avançado estágio da MDS, 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 nos estágios anteriores.

NA PRÁTICA...

Toda data stack construída com base nas operações básicas descritas 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.

E todos esses conceitos abordados até aqui serão úteis à medida que avançamos nas explicações sobre as ferramentas e tecnologias recomendadas para a implementação de uma **Modern Data Stack** nos seus negócios.



6 | Data stack: principais ferramentas da abordagem moderna de dados

Uma das características mais marcantes da **Modern Data Stack** é a junção de diversas ferramentas e tecnologias para responder demandas das operações de dados atuais.

Por isso, além dos estágios do fluxo de dados, as ferramentas usadas em cada etapa da **data stack** são elementos essenciais que determinam o sucesso da implementação da **MDS**.

Uma operação de dados moderna requer, portanto, a combinação de diversos serviços e ferramentas em uma **data stack**.

Confira a análise da **Indicium** sobre essas principais ferramentas disponíveis no mercado para a construção de uma abordagem moderna de dados, escalável e flexível na sua empresa.



6.1 | Ferramentas de coleta, deployment e transformação de dados

Coleta e integração de dados

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



Deployment

Docker e **Kubernetes** são ferramentas muito utilizadas para realizar o deployment em conjunto com orquestradores, como **Airflow** e **Prefect**. O diferencial é que elas fazem 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

O **dbt** (data build tool), o **Dataform** e o **Spark** são as três principais ferramentas da **Modern Data Stack** para a transformação de dados. Juntas, elas 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 dados.



6.2 | Ferramentas de armazenamento e gerenciamento de dados

Hoje, temos duas opções práticas, viáveis e eficientes quanto ao armazenamento de dados, tanto on premises quanto em nuvem: **data warehouse** e **data lake**. Ambas devem ser avaliadas caso a caso, pois apresentam diferenças técnicas e conceituais em termos de arquitetura e finalidade.

Data warehouse

O **Amazon Redshift**, **Snowflake**, **Google Big Query** e **Post Gre SQL** têm arquitetura escalável em nuvem, permitindo o armazenamento e a consulta de volumes enormes de dados de forma rápida. Por isso, essas ferramentas são essenciais na construção de uma **abordagem moderna de dados**.

Data lake

O **Dremio** e o **Amazon S3** são as mais indicadas porque armazenam em um só lugar todos os tipos de dados, estruturados, não estruturados e híbridos.

Computação em nuvem

A **AWS**, o **Google Cloud** e **Microsoft Azure** são hoje os principais provedores de computação em nuvem. É possível escolher entre essas opções de acordo com a necessidade da sua empresa para o armazenamento e gerenciamento dos seus dados.





6.3 | Ferramentas de data analytics

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

Business intelligence

O **Metabase** é destaque entre diversas alternativas robustas e open source; e o **Microsoft PowerBI**, **Looker** e **Tableau** destacam-se como plataformas SaaS (software as a service, ou software como serviço). 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.

Advanced analytics

O **MLFlow** e o **Kedro** auxiliam na execução de **modelos preditivos e prescritivos**, e otimizam o processo de desenvolvimento, reduzindo o tempo entre a modelagem e a utilização do usuário final. **Machine learning, inteligência artificial e modelagem de dados** são técnicas aplicadas em **advanced analytics** para análises mais complexas dentro da **data stack** por meio dessas ferramentas e das diversas bibliotecas nas linguagens **R** e **Python**.



Metabase



Power BI



Looker



+tableau



mlflow



Kedro



7 | Conclusão

São muitas as decisões envolvidas no desenvolvimento e aprimoramento da operação de dados de uma empresa.

Informações atualizadas podem ser difíceis de encontrar e ainda mais difíceis de entender.

A **Modern Data Stack (MDS)** soluciona esse problema criando uma **metodologia integrada de inteligência de negócios** que coleta, combina, analisa e entrega o valor dos dados para os negócios.

Felizmente, os componentes da **data stack** estão muito mais baratos, simples de configurar e manusear. Assim, empresas de todos os portes podem utilizá-la 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.

Entre em contato conosco



CONHEÇA A

indiciu



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 infraestruturas, a elaboração de indicadores e o desenvolvimento de modelos preditivos, possibilitando o **fácil acesso à informação para tomada de decisões em tempo real**.



Top B2B Services Clutch Brazil 2022

A Indicium foi eleita a **terceira entre as top companhias de data analytics da América Latina**. O ranking foi realizado pela **Clutch**, plataforma líder mundial em pesquisa, avaliação e reviews para compradores e contratantes de serviços B2B.

Esse reconhecimento é fruto do **trabalho sério e comprometido da nossa equipe** de projetar, criar e implementar **projetos de data analytics e business intelligence de excelência**.



Top Big Data Analytics Company DesignRush 2022

A Indicium foi reconhecida como **uma das principais empresas de IA em 2021 pela DesignRush**.

Temos a honra de ser reconhecidos entre as **30 principais agências de inteligência artificial**. Isso reforça nosso compromisso de aumentar o ecossistema de dados na América Latina e oferecer **melhores decisões de negócios para nossos clientes com nossas soluções de dados exclusivas**.



Ranking EXAME Negócios em Expansão

A **Indicium** é uma das maiores empresas de dados do Brasil e da América Latina, e **ficou em sexto lugar no Ranking EXAME Negócios em Expansão 2022.**

Somos uma **grande referência** enquanto empresa de dados, internacionalmente reconhecida nas nossas **áreas de data science e analytics.**



dbt Labs Platinum 2022 Certification Award Winner

A **Coalesce 2022 by dbt Labs** é um grande encontro da comunidade de dados mundial, e este ano a **Indicium** volta de lá como empresa de dados **líder em certificações.**

Ganhamos status de parceria Platinum por ter certificado o maior número de profissionais nas qualificações disponíveis no evento.



Innoway Acate

A **Indicium** recebeu o **selo Cultura de Inovação Avançada - Innoway Acate**, que certifica empresas capazes de impactar positivamente o mercado e a sociedade.

Esse projeto dá destaque e protagonismo para as pessoas enquanto responsáveis por inovar dentro das empresas, buscando conscientizar as gestões organizacionais para a necessidade da adoção de metodologias cada vez mais humanizadas. A **cultura de inovação** está no nosso DNA.



Great Place To Work

A **Indicium** recebeu o selo **Great Place To Work (GPTW)**, que certifica empresas capazes de se adaptar e **inovar com agilidade organizacional**, e que promovem **experiências positivas e todas as condições para a inovação** acontecer de dentro para fora.

É uma honra receber este selo, pois se trata de uma avaliação interna, um termômetro do que vivemos aqui enquanto criamos soluções de dados inovadoras ao mesmo tempo em que formamos pessoas responsáveis por inovar.

www.indicium.tech



indiciumtech



indiciumtech



indicium_tech



indiciumtech

Rua Orlando Phillippi, n. 100, Edifício TechPlan, 3º andar, sala 304
Saco Grande - Florianópolis - SC - Brasil CEP 88032-700
Fone: +55 (48) 3371-7293