

alteryx

Self-service analytics no Google Cloud Platform:

5 lições de preparação de dados para garantir o sucesso

Você deseja solucionar desafios analíticos por conta própria e impulsionar a produtividade e a inovação, mas necessita de dados adequados.

Diante da complexidade do cenário atual, aplicativos tradicionais de business intelligence e armazenamentos de dados deixaram de ser suficientes. Mesmo quando é possível desbloquear data warehouses com ferramentas típicas como o Tableau ou o Qlik, ainda é necessário recorrer ao uso de planilhas e ao suporte da TI.

A resposta é uma solução de self-service analytics.

Neste documento, descrevemos cinco lições aprendidas ao ajudar empresas na migração para o Google Cloud Platform, uma suíte analítica inteligente e integrada, que oferece flexibilidade para lidar com todos os tipos de dados, em qualquer escala ou velocidade.





Lição nº 1

Eliminar o modelo de migração "lift and shift"


Muitas organizações adotam a abordagem "lift and shift" — também conhecida como rehosting, ao migrar aplicativos para a nuvem. Essa estratégia não envolve mudanças na arquitetura e praticamente nenhuma alteração no código. Apesar de facilitar a gestão das cargas de trabalho, esse modelo é incapaz de aproveitar o potencial da nuvem para aprimorar resultados analíticos, promover benefícios nos negócios ou reduzir custos operacionais. O processo e as ferramentas continuam iguais, impossibilitando usuários de perceber o valor efetivo de um ambiente self-service.

Para ser uma solução verdadeiramente eficaz, é fundamental incluir aplicativos nativos na nuvem. Além disso, processos de coleta, armazenamento e tratamento de dados, assim como a geração de relatórios, devem ser projetados para permitir a integração com sistemas desenvolvidos exclusivamente para a nuvem.

Ela também deve oferecer suporte a ambientes em nuvem que sejam dinâmicos, flexíveis e escalonáveis, incluindo contêineres e ofertas de microsserviços. Ao invés de simplesmente recriar plataformas monolíticas no local, organizações bem-sucedidas aproveitam a elasticidade da nuvem e softwares de código aberto,

como Kubernetes ou Docker, para fazer a orquestração de contêineres e fluxos de dados.

O Google Cloud Platform é a solução ideal para implementar o self-service analytics porque oferece uma arquitetura sem servidor, incluindo serviços de inteligência analítica. Você pode ativar cada componente com facilidade, escalar de forma dinâmica ou reduzir a alocação de recursos com base no uso. Operações de manutenção não precisam ser planejadas. Com a capacidade de alavancar diversas funções analíticas, aproveitar modelos flexíveis e controlar custos, organizações conseguem focar no valor dos dados para impulsionar os negócios.



Lição nº 2 ETL e self-service são incompatíveis

Tecnologias herdadas de integração de dados, como ETL (extrair, transformar e carregar), podem ser aplicadas na criação de pipelines para transferir dados armazenados no local e, em outras nuvens, para o Google Cloud Platform. Por outro lado, aplicativos modernos de integração e streaming de dados, tais como StreamSets, Fivetran, ou Google Cloud Data Fusion, são eficientes para facilitar fluxos de dados entre sistemas e, torná-los disponíveis para a análise no Google Cloud Platform.


Após construir um data lake ou data warehouse, pode ser tentador continuar usando o processo de ETL, mas ele é simplesmente inadequado quando se trata de self-service analytics.

Normalmente, envolve atividades técnicas e complexas destinadas a engenheiros de dados, que trabalham com transformações recorrentes em tempo real e em lotes definidos para movimentar dados em larga escala. Para obter acesso e aprofundar a análise, equipes especializadas precisam executar tarefas adicionais para preparar as informações.

Como parte da suíte analítica inteligente, o Google Cloud Platform oferece uma solução completa para o preparo de dados: o Cloud Dataprep by Trifacta. Assim,

usuários podem avaliar a qualidade, refinar e padronizar as informações, assim como fazer a combinação e o cálculo de múltiplos dados.

O Cloud Dataprep by Trifacta foi projetado para auxiliar todos os profissionais do setor — engenheiros de dados, analistas, usuários corporativos e demais equipes analíticas — a desbloquear data lakes ou data warehouses ao interagir com o conteúdo de dados para refiná-los de forma iterativa, e agrupá-los para abastecer análises posteriores.



Lição nº 3 Self-service não significa livre para todos

Quando mais pessoas conseguem aplicar recursos de self-service analytics, o número de usuários pode aumentar. Mas isso não significa que as informações são livres para todos. Habilitar novas equipes com ferramentas self-service exige uma governança rigorosa para:

- Evitar a proliferação de dados fora de controle
- Cumprir exigências regulatórias
- Manter a confiabilidade dos dados para a tomada de decisões comerciais

É importante estabelecer o equilíbrio ideal entre proteger ativos de dados e aumentar a colaboração de usuários para extrair valor das informações, conforme acharem pertinente.



Organizações precisam encontrar o equilíbrio ideal entre proteger ativos de dados (sob perspectivas de governança e segurança), e aumentar a colaboração de usuários para extrair valor das informações, conforme acharem pertinente. Existem três maneiras de alcançar esse equilíbrio.

Reduzir silos

Quando o armazenamento e processamento de dados são centralizados na nuvem com escalabilidade praticamente ilimitada, incluindo a opção do usuário final introduzir seus próprios dados, é possível evitar a proliferação de silos de dados. Com isso, equipes podem coletar fragmentos de dados, executar tarefas específicas de preparação, além de criar relatórios dentro e fora da nuvem, ao invés de extrair informações redundantes e duplicá-las em planilhas eletrônicas.

Utilizar um catálogo de dados

Recursos compartilhados, como catálogos ou glossários para gerenciar definições de dados, assim como metadados e conhecimentos sobre a linhagem de dados, permitem encontrar informações com mais rapidez, além de auxiliar organizações a gerenciar fontes de dados e monitorar todo o ciclo de vida. Soluções de machine learning (ML) e inteligência artificial (IA) automatizam a coleta e gestão de metadados, contribuindo para gerar insights relevantes.

Rastrear e documentar a linhagem de dados

A linhagem de dados — ou seja, como os dados foram usados e transformados, além da pessoa responsável — é um fator imprescindível para gerar relatórios e auditorias regulamentares. Ademais, é fundamental para aqueles encarregados na tomada de decisões, que precisam entender o histórico das informações envolvidas no processo analítico, visualizações e recomendações prescritivas, assim como o impacto de novos requisitos sobre o pipeline de dados para executar a análise.

Lição nº 4

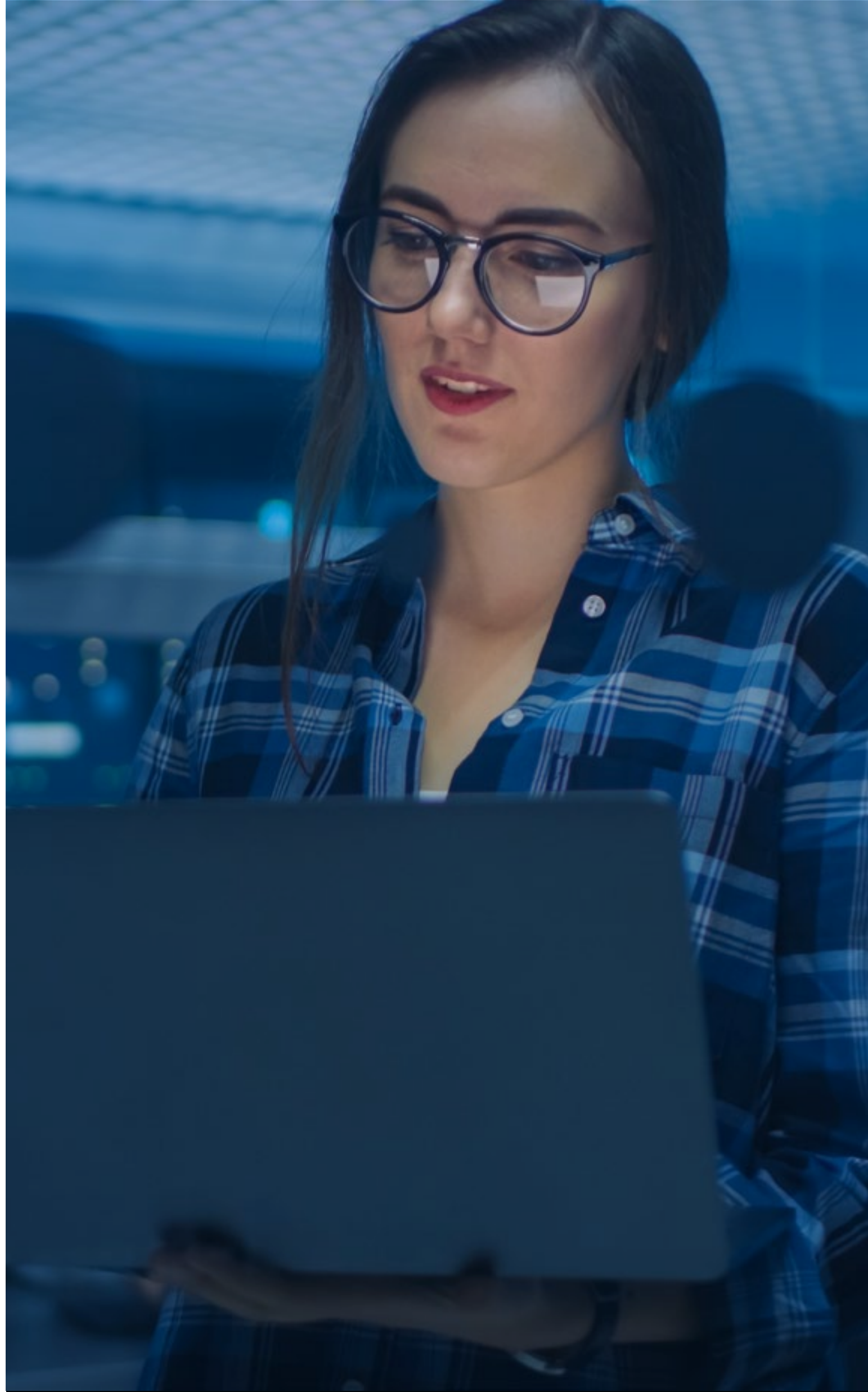
Reavaliar funções e atribuições

O preparo de dados self-service sugere que organizações reavaliem funções e atribuições das principais partes incluídas na cadeia de valor e, possam reimaginar processos rígidos envolvendo pipelines de dados, como fluxos mais ágeis e dinâmicos. Cada parte interessada aproveita o preparo de dados self-service de diferentes maneiras para atender suas necessidades e, ao mesmo tempo, incorpora uma estrutura compartilhada de projetos, comunicação e colaboração para aprimorar análises.

O preparo de dados self-service introduz novas abordagens para aumentar a colaboração entre engenheiros de dados, analistas, data scientists, arquitetos de dados, além de executivos e gerentes de negócios focados na análise.

ANALISTAS DE DADOS, assim como gestores corporativos, gerentes de projetos e funções relacionadas, normalmente trabalham de perto com usuários de negócios. Muitas vezes, exploram dados brutos a fim de descobrir aspectos relevantes para responder perguntas específicas. O principal objetivo é obter respostas rápidas com a maior facilidade possível.

O **Cloud Dataprep by Trifacta** é ideal para analistas de dados porque é simples de implementar e dispensa conhecimentos especializados de codificação. Equipes podem interagir desde o início do processo de preparação, bem como explorar e modelar protótipos para se adaptar às demandas dos negócios. Quando a automação é necessária para assegurar sustentabilidade e repetibilidade, todos podem colaborar com engenheiros de dados para orquestrar o pipeline de ponta a ponta.



ENGENHEIROS DE DADOS projetam, constroem e gerenciam o processamento e a arquitetura de dados para apoiar projetos analíticos. Eles participam ativamente da transformação de dados, incluindo a exploração e o perfil de dados brutos. A intenção desses profissionais é otimizar e automatizar um grande volume de processos de dados para gerenciá-los.

O **Cloud Dataprep by Trifacta** é perfeito para engenheiros de dados porque contribui para operacionalizar e monitorar diversos fluxos de dados, projetados tanto por eles, quanto por analistas. Ele facilita a colaboração entre equipes especializadas e todas as partes envolvidas, para entender requisitos de infraestrutura e guiar usuários sobre novas formas de explorar, analisar, modelar e consumir dados.

CIENTISTAS DE DADOS aplicam conhecimentos e habilidades técnicas para projetar e modelar algoritmos, potencializando recursos de machine learning e inteligência artificial. No entanto, podem gastar até 80% do tempo com o preparo de dados, limitando iniciativas de inovação.

O **Cloud Dataprep by Trifacta** é adequado para cientistas de dados porque é uma solução nativa em nuvem, que simplifica tarefas habituais na preparação de dados, possibilitando a delegação de recursos com maior disponibilidade e menos custos.

Lição nº 5

A qualidade de dados é responsabilidade de todos

Todas as partes interessadas precisam saber que podem confiar nos dados que impulsionam análises, visualizações e algoritmos. O resultado de qualquer modelo preditivo ou insight analítico tem a mesma qualidade dos dados envolvidos.

Todas as organizações devem se comprometer em priorizar a limpeza de dados.

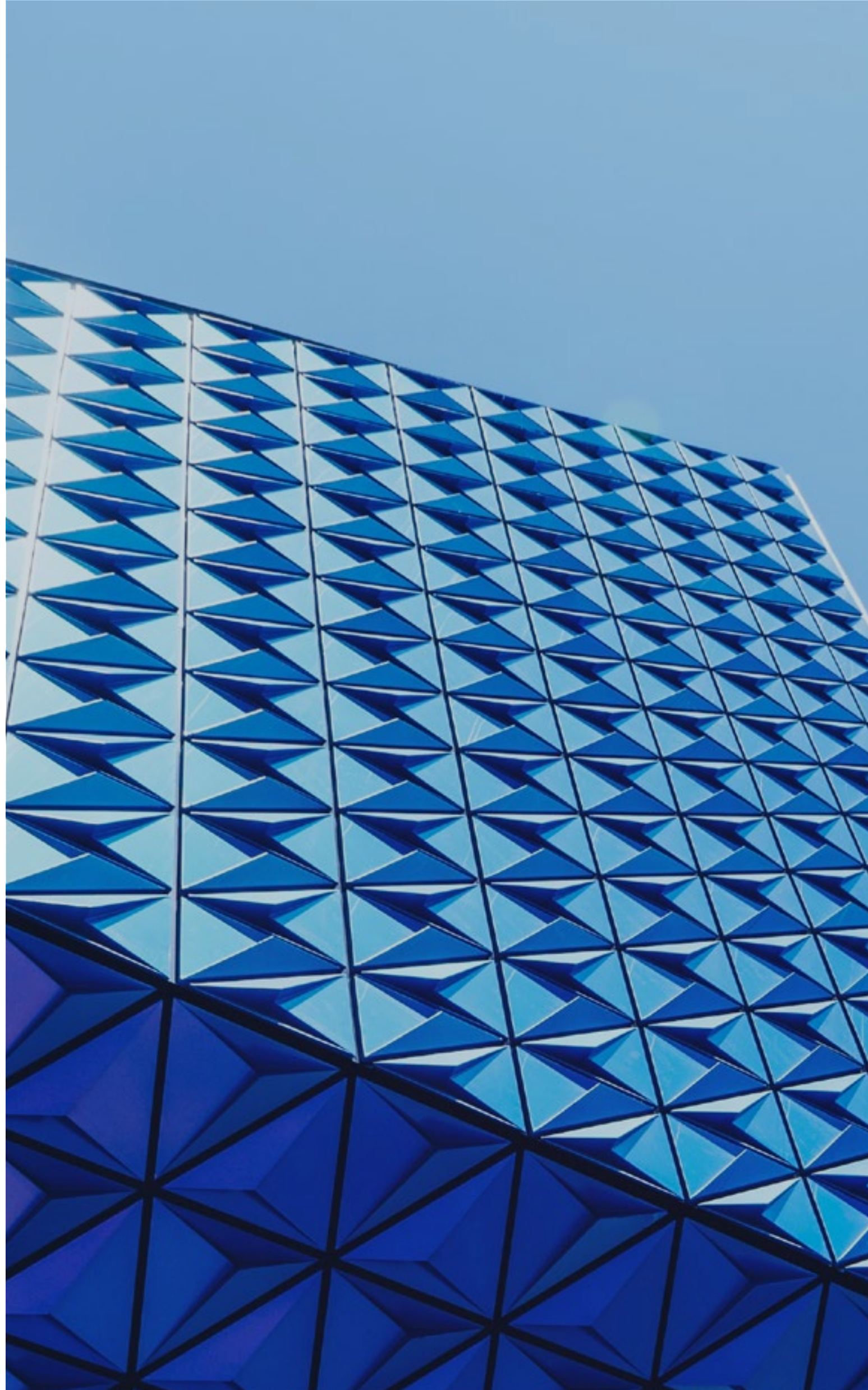
Embora o ideal seria tratar a qualidade de dados em nível corporativo, isso é insustentável. O volume de dados e a enorme variedade de fontes (internas e terceirizadas), além dos diferentes tipos e contextos de dados, dificultam avaliar a integridade, corrigir falhas e monitorar novas informações para identificar possíveis problemas. Repositórios de dados modernos, como data lakes e soluções de armazenamento em nuvem, apresentam desafios ainda mais complexos.

Tratar a qualidade de dados na fase de preparação é mais eficaz. O Cloud Dataprep by Trifacta melhora a precisão, a consistência e a confiabilidade dos dados, aplicando recursos de ML e IA para automatizar processos de limpeza. A automação controla a escala de grandes armazenamentos de dados e identifica rapidamente valores que aparentam estar incorretos, inválidos, ausentes ou inadequados. Anomalias que justificam uma investigação mais detalhada são automaticamente sinalizadas.

A automação de perfis de dados e processos de limpeza detectam inconsistências entre as fontes

integradas ao banco de dados, além de apontar redundâncias e recomendações visuais para solucionar questões de qualidade com transformações automatizadas sem código.

Conforme novos dados estruturados, não estruturados ou semi-estruturados são introduzidos e integrados na nuvem, a qualidade é validada constantemente. "Validação contínua" significa que usuários não precisam esperar a conclusão do processo para visualizar e testar os resultados, um fator incompatível com as atuais metodologias de desenvolvimento ágeis.



Conclusão

A Alteryx contribuiu com o sucesso de inúmeras organizações para a implantação do self-service analytics na nuvem. Com essa experiência, aprendemos cinco lições principais:

Nº 1. Usar uma simples abordagem "lift-and-shift" para migrar o analytics local para a nuvem não funciona.

Selecione uma plataforma em nuvem com componentes analíticos projetados especificamente para a nuvem, como o Google Cloud Platform, que apresenta um conjunto analítico inteligente e um ambiente sem servidor

Nº 2. As ferramentas de ETL podem suavizar o atrito entre a transferência de dados locais para repositórios em nuvem, mas são inflexíveis para recursos de self-service analytics. A melhor alternativa é adotar soluções modernas de preparação de dados, como o Cloud Dataprep by Trifacta, capaz de habilitar usuários com qualquer nível de habilidade a aproveitar dados relevantes para a análise.

Nº 3. Self-service não é um ambiente sem regras, livre para todos. Para estabelecer uma estrutura eficaz, organizações devem encontrar o equilíbrio ideal entre a governança e a disponibilidade.

Compreender a linhagem de dados e desenvolver um glossário de negócios é determinante para o sucesso.

Nº 4. Adotar processos ágeis para enfatizar a colaboração, assim como uma linguagem comum entre todas as partes interessadas, para transformar dados em ativos valiosos para a análise.

Nº 5. A qualidade de dados é responsabilidade de todos. O Cloud Dataprep by Trifacta avalia automaticamente a qualidade dos dados, além da validação e resolução contínuas, para garantir a confiabilidade das informações integradas ao self-service analytics.

alteryx

SOBRE A ALTERYX

A Alteryx, empresa líder em automação analítica, tem como objetivo potencializar cada pessoa a transformar dados em descobertas. Oferecemos uma plataforma unificada para automatizar processos de negócios, analytics e data science, permitindo acelerar a transformação e moldar o futuro digital. Organizações de todos os portes ao redor do mundo confiam na plataforma Alteryx Analytics Automation para alcançar resultados de alto impacto e impulsionar a qualificação da força de trabalho. Para saber mais, visite www.alteryx.com/pt-br.

Alteryx é uma marca registrada da Alteryx, Inc. Todos os outros produtos e nomes de marcas podem ser marcas comerciais ou registradas de seus respectivos proprietários.

Alteryx Analytics Cloud

Conecte-se, crie perfis, transforme e compartilhe dados através de pipelines automatizados com o Alteryx Analytics Cloud. Com pouco ou nenhum código, oferecemos tudo que você precisa. Explore visualmente a qualidade dos dados, acelere o preparo para análises e, crie e implante pipelines em segundos.

[SAIBA MAIS >](#)

Alteryx Machine Learning

Dimensione processos de data science em toda a empresa com o poder do machine learning automatizado (AutoML) e a engenharia de recursos, qualificando tanto especialistas no domínio de negócios, quanto cientistas de dados, a acelerar a produção de insights.

[SAIBA MAIS >](#)

Alteryx Designer

Automatize cada etapa da análise, incluindo processos de preparo e combinação de dados, geração de relatórios, análise preditiva e projetos de data science. Acesse qualquer fonte, arquivo, aplicativo ou tipo de dados, e experimente a simplicidade e o poder de uma plataforma self-service com mais de 300 blocos de automação. Comece a criar resultados interativos hoje mesmo.

[SAIBA MAIS >](#)

Alteryx Intelligence Suite

Seja você iniciante ou especialista, crie modelos revolucionários de machine learning em poucos minutos. O Alteryx Intelligence Suite foi projetado para ajudar você a desbloquear insights ocultos nos dados. Desde a análise de dados estruturados até a descoberta de insights inéditos em textos e documentos, o Intelligence Suite contribui para solucionar os principais desafios do negócio.

[SAIBA MAIS >](#)

Alteryx Server

Dimensione, compartilhe e controle processos analíticos, modelos e fluxos de trabalho com o Alteryx Server. Crie fluxos de trabalho analíticos com o Alteryx Designer e publique no Alteryx Server para agendar relatórios e resultados automatizados. A governança integrada, a segurança centralizada e a alta disponibilidade, auxiliam empresas a ampliar a escala da análise em toda a estrutura organizacional.

[SAIBA MAIS >](#)