

A empresa impulsionada por dados

Por Mark Schwartz, Estrategista corporativo, AWS





Introdução

Hoje, há muitos debates sobre a empresa impulsionada por dados e a urgência de tornar-se uma. Mas, o que exatamente configura essa condição de ser impulsionada por dados e por que isso é tão importante para o ambiente digital atual? Que medidas práticas uma empresa deve tomar para tornar os dados fundamentais para seus parâmetros e práticas? E qual é a conexão entre dados e a famosa prioridade da era digital: a agilidade técnica e empresarial? Este eBook apresentará o que é ser impulsionado por dados e alguns exemplos de como as empresas estão usando dados para conduzir seus negócios. Também faremos uma conexão de pontos entre o que é tornar-se impulsionado por dados e a agilidade, a transformação digital e a inovação contínua.

Empresas impulsionadas por dados se esforçam para ***fundamentar suas decisões estratégicas de negócios*** em indícios fornecidos por dados, o que exige um nível de rigor e, ao mesmo tempo, uma habilidade para inovar e identificar nos dados oportunidades que possam originar novos produtos ou mercados. Elas também surgem para trabalhar com dados como um ativo que pode ser usado tanto para melhorar as interações com clientes, como para aumentar a eficiência. Em outras palavras, elas ***analisam*** dados para orientar a tomada de decisões e os ***utilizam*** para atender aos seus clientes. Dados podem funcionar como a base, por exemplo, para personalização, definição de preço dinâmica, expansão de mercado, inovação de produtos ou otimização da cadeia de suprimentos.

Contudo, até pouco tempo, algumas empresas achavam difícil utilizar dados para esses fins, pois elas apenas consideravam seu uso em contextos de transações. Consequentemente, elas os armazenavam a sete chaves em bancos de dados isolados que eram excelentes para processamentos de transações, porém menos adequados para análises abertas. Nosso modelo mental era o mesmo da fatura ou do formulário do pedido: “Gostaria de 20 widgets por US\$ 100 cada”. Ou, “Me pague por 20 widgets, cada um por US\$ 100”. Os dados eram performativos e imperativos, um estímulo ou elemento para conduzir transações. Hoje, o valor dos dados vai muito além de sua função transacional.

Como podemos pensar nesse valor em termos financeiros e como podemos maximizá-lo?

O valor empresarial dos dados

Cada parcela de dados pode ser usada em qualquer quantidade de análises que possam originar resultados empresariais. Logo, o valor também reside em possibilitar resultados a partir de análises. Por exemplo, se a empresa analisa seu histórico de transações e, como resultado, descobre maneiras de otimizar a cadeia de suprimento (conseguindo reduzir custos), então os dados já desempenham uma função na redução de gastos. Assim, os dados adquirem um valor de negócios derivado de seu uso potencial no aumento de lucros ou na conquista de objetivos de missão crítica.

É fácil encontrar instâncias de dados sendo utilizadas por seu valor não transacional. A Johnson & Johnson, por exemplo, utiliza os dados transacionais que armazenou na nuvem para melhorar a conformidade física, otimizar sua cadeia de suprimento e descobrir novos medicamentos. A Nike coleta dados sobre conquistas de clientes para nortear a experiência digital do cliente na NikePlus. A Lyft coleta e armazena as coordenadas de GPS de todas as suas corridas; ao analisá-las, descobriram que 90% das corridas estavam sobrepostas com outras corridas de localidades próximas. Essa descoberta levou à criação do Lyft Line, um serviço que permite que passageiros compartilhem carros e recebam descontos de até 50%¹.

Como esses usos podem originar futuros lucros, ainda que não percebidos, podemos considerar dados como ativos financeiros (embora sejam ativos não vinculados à GAAP, na maioria dos casos). Não é nenhuma surpresa que, por esta razão, os dados acumulados por uma empresa se tornem um fator determinante no valor de aquisição da própria empresa ou possam capacitá-la a formar parcerias com outras iniciativas corporativas. Um bom exemplo disso é a aquisição do LinkedIn pela Microsoft, com dados de 433 milhões de clientes, por US\$ 26,2 bilhões, ou os processos de falência da Caesars Entertainment Operating Corp. Inc. de 2015 a 2017, em que credores alegaram que os dados de 45 milhões de clientes presentes em seu programa de fidelidade Total Rewards valiam US\$ 1 bilhão e eram o ativo mais valioso da empresa².

Facilita pensarmos em dados como detentores de valor corporativo como uma espécie de opção de chamada financeira, ou seja, eles nos dão a oportunidade de mudar a cadeia de suprimentos ou lançar um novo produto, sem que isso seja uma obrigação. Podemos exercitar tal opção ou não, dependendo da indicação de valor oferecida pelos dados a respeito de um determinado novo empreendimento. É aqui que encontramos uma certa dificuldade em identificar o valor de dados como ativos: avaliar o valor de uma opção de chamada é bem mais complicado do que calcular o ROI da circulação projetada de fluxos de caixa. Como resultado, algumas empresas frequentemente negligenciam esse valor; mas, conforme mostro no meu livro *War and Peace and IT*³, muitas das tecnologias de entrega ágil de TI resultam nesse tipo de valor opcional.

¹ Estudos de caso da AWS. Consulte <https://www.youtube.com/watch?v=6A1tOFqvgek>, <https://aws.amazon.com/products/databases/> e <https://aws.amazon.com/solutions/case-studies/lyft/>.

² Os dois exemplos do <https://sloanreview.mit.edu/article/whats-your-data-worth/>. Uma análise detalhada da falência da Caesars pode ser encontrada em <https://turnaround.org/sites/default/files/11.%20Paper%20-%20Caesars.pdf>. O processo de falência foi excessivamente complexo e o valor do Total Rewards foi incluído em outros ativos, de forma que não ficou claro qual valor ficou atrelado a ele no final.

³ Mark Schwartz, *War and Peace and IT: Business Leadership, Technology, and Success in the Digital Age* (Portland, OR: IT Revolution Press, 2019).



Dados e agilidade

O valor é criado não só em função dos dados em si, mas também a partir das ferramentas e processos em vigor para analisá-los e produzir os resultados de negócios. No mundo digital atual, carregado por mudanças aceleradas, insegurança e complexidades, que talvez possam ser resumidas a desestruturação, precisamos usar dados para dar suporte à agilidade empresarial e responder de forma rápida e flexível às circunstâncias em contínuo processo de adaptação. A agilidade é o que permite que organizações transformem as mudanças constantes em oportunidades e evitem a desestruturação, respondendo de forma ágil às ameaças competitivas. As empresas da era digital aprenderam que precisam inserir as versões mais primárias dos produtos no mercado e desenvolvê-las por meio de feedback contínuo do próprio mercado⁴.

Os últimos anos proporcionaram técnicas para implementar agilidade no **processo** de desenvolvimento do produto, incluindo o desenvolvimento de software ágil, DevOps e o desenvolvimento de software lean. A nuvem tem sido usada para acelerar a entrega de capacidades de TI, tanto para software, como hardware. Estruturas organizacionais baseadas em equipes possibilitaram a mobilização de recursos para atender às necessidades em constante transformação. Todos esses desenvolvimentos ajudaram empresas a tornar seus processos internos mais ágeis.

Mas processos ágeis representam uma parcela da história: os próprios dados da empresa também precisam ser ágeis. É necessário que eles estejam facilmente disponíveis para usos inesperados e em constante mudança. Eles precisam ser acessíveis e pertinentes. Os funcionários devem ter ferramentas facilmente disponíveis para trabalhar com dados, bem como as habilidades necessárias para fazer isso. Essa habilidade de usar a flexibilidade dos dados para torná-los disponíveis para usos que não temos conhecimento previamente representa o elo perdido da conquista da agilidade empresarial, é o que diferencia organizações ágeis de outras que meramente adotam as estruturas e as armadilhas dos modelos ágeis. Agilidade empresarial requer agilidade dos dados. Uma empresa impulsionada por dados domina as duas.

Esse foco na conquista de agilidade para dados é inédito. Como os dados eram apenas transacionais, nós podíamos mantê-los isolados em bancos de dados altamente estruturados, cuja estrutura refletia a forma como seriam usados por tais transações. Nossas ferramentas eram sistemas de bancos de dados relacionais, como Oracle ou SQL Server, cujo forte está no processamento transacional. Usávamos dados para conduzir as transações em si e produzir relatórios operacionais para dar suporte às transações.

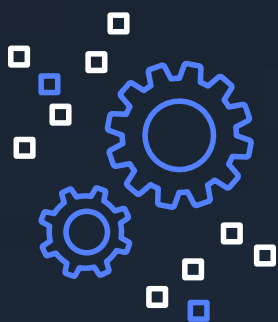
Na medida em que prestávamos atenção à privacidade, nós a garantíamos limitando o acesso aos dados de forma rigorosa, ao invés de buscar formas de disponibilizá-los nos limites das defesas da privacidade. Em vez de usarmos a privacidade de acordo com o projeto, nós exercíamos uma espécie de privacidade por ocultação.

Sim, nós tivemos tentativas de liberar dados para análises ad hoc com os chamados sistemas de inteligência de negócios (BI). Mas hoje as ferramentas evoluíram para muito além do que os sistemas BI podem alcançar: nós agora temos machine learning, uma variedade de bancos de dados criados especificamente para lidar com diferentes tipos de dados, algoritmos para processamento massivamente paralelo, grandes quantidades de dados não estruturados, como vídeo e fala, dispositivos IoT que fornecem fluxos de dados derivados de sensores, bem como grandes quantidades de dados. Com tais ferramentas, nós podemos liberar nossos dados de seu contexto transacional e operacional.

Acima de tudo, percebemos que ser impulsionado por dados não é apenas um desafio técnico, mas também organizacional. **Para ser impulsionada por dados, uma empresa deve pensar de forma distinta sobre como tomar decisões de negócios e como interagir com seus clientes. Trata-se de um compromisso com o valor de dados, um tipo de humildade empresarial que admite os dados sabem mais do que nós.**

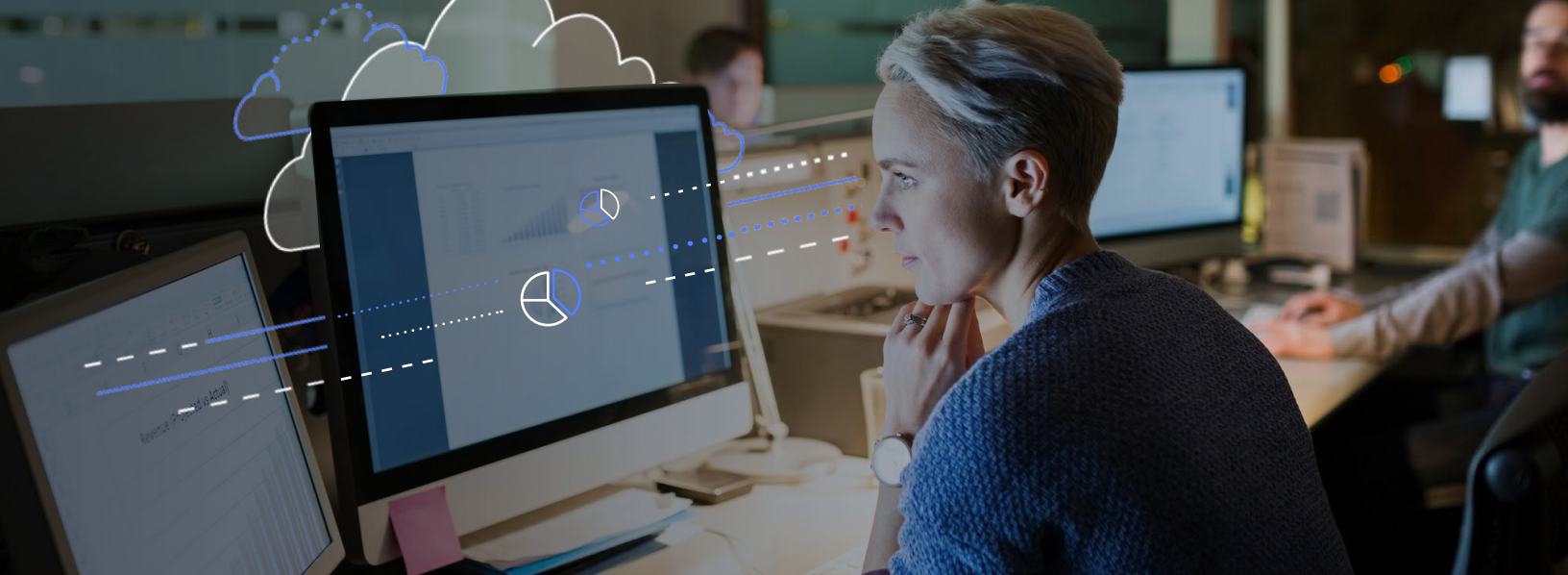
Como podemos disponibilizar nossos dados para serem usados de formas inesperadas; ou seja, como podemos nos beneficiar de sua flexibilidade para conquistar agilidade empresarial? Como podemos aplicá-los para proporcionar rigor e criatividade na tomada de decisão de negócios? Como podemos mudar a cultura de negócios para nos beneficiar dessa nova flexibilidade?

E como podemos aplicar defesas de controle adequadas em torno de dados para proteger sua privacidade ao permitirmos seu uso com flexibilidade e agilidade?



Duas perguntas importam:

- 1. Como podemos implementar agilidade em nossos dados?**
- 2. Como podemos usar dados para implementar agilidade nos negócios?**



Agilidade nos dados

Como podemos implementar agilidade em nossos dados?

Para conquistar agilidade nos negócios, nós precisaremos estar prontos para responder a mudanças inesperadas nos negócios e ambientes competitivos. Também precisaremos criar inovações verdadeiramente inéditas e, dessa forma, saber colocar nossos dados em ação de maneira que não saberemos necessariamente quando os coletarmos.

Nossos desafios:

- É possível que nossos dados estejam armazenados a sete chaves em bancos de dados relacionais e transacionais, provavelmente isolados de formas que os deixam inacessíveis a diferentes partes da nossa empresa.
- Pode ser que nós não tenhamos as ferramentas analíticas adequadas ou elas não estejam disponíveis para as pessoas certas nos momentos certos.
- Nossos modelos de segurança e privacidade são ad hoc, como talvez nunca os tenhamos contemplado usando dados para explorar. O mais provável é que estejamos promovendo a privacidade apenas por manter nossos dados o mais inacessível possível.

Nossas metas:

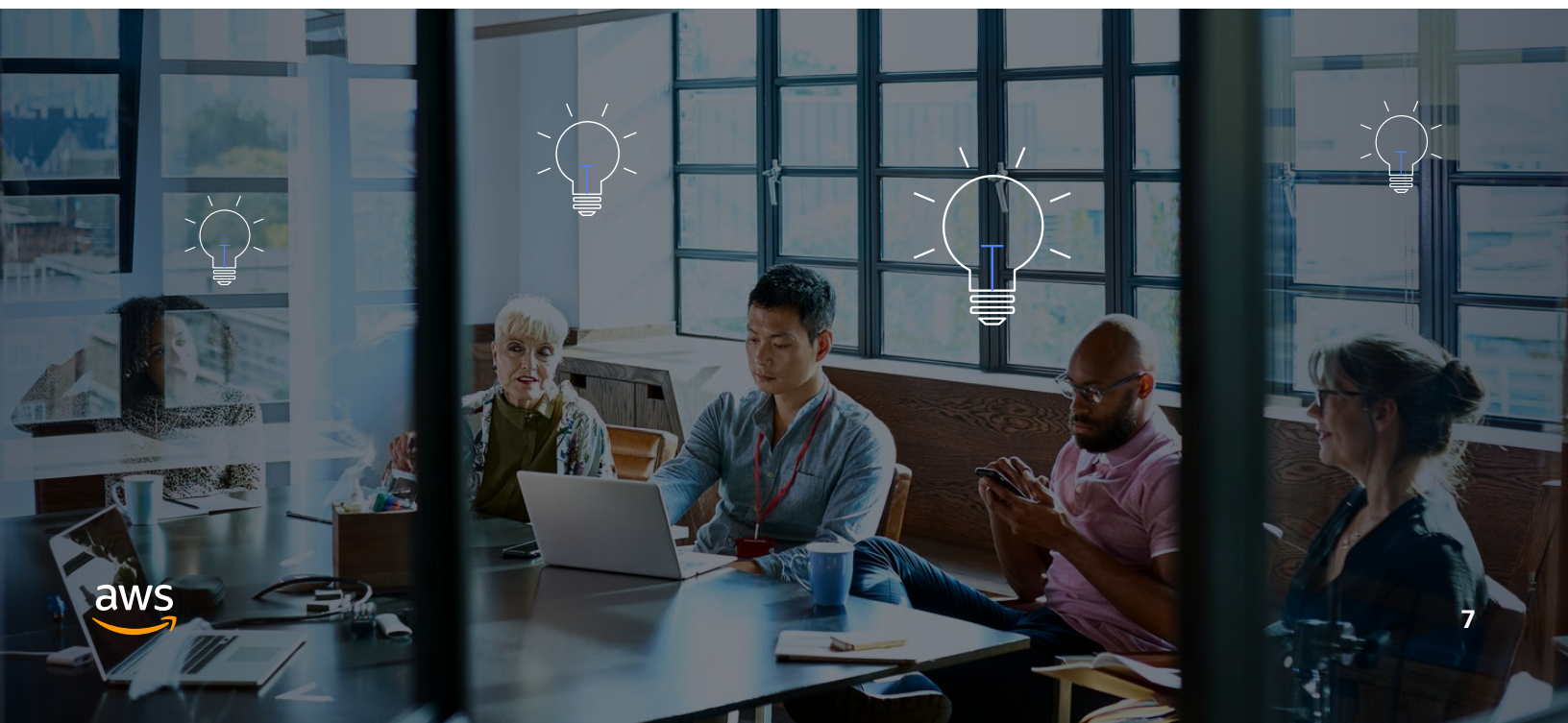
- Maximizar a disponibilidade dos dados sem perder de vista as proteções à privacidade e confidencialidade.
- Promover transparência em toda a empresa derrubando silos de informações.
- Oferecer aos funcionários as ferramentas adequadas para explorar dados de maneiras não planejadas e aproveitando os últimos avanços em análises.
- Certifique-se de que tem conhecimentos necessários para interpretar dados de maneira rigorosa e criativa.

No livro [*"Analytics without Limits: FINRA's Scalable and Secure Big Data Architecture"*](#), John Brady, diretor executivo de segurança da informação da Autoridade Regulatória do Setor Financeiro (FINRA), enquadra tais objetivos de forma elegante dizendo que deseja **reduzir o custo da curiosidade**. Ele se refere a custos em sua forma mais ampla, incluindo o tempo gasto para delinear inferências de dados e o risco de disponibilizá-los. O negócio da FINRA é explorar as 37 bilhões (ou mais) transações em vigor no mercado financeiro diariamente, buscando padrões de fraudes. Como nem sempre ela sabe de forma antecipada como um padrão de fraude se dá, muitas vezes seus funcionários precisam confiar nos conhecimentos de seus analistas para localizar comportamentos suspeitos. Sua tarefa tem total conexão com a curiosidade: ela quer que seus analistas examinem dados de forma curiosa para identificar padrões existentes e o porquê. A tarefa de sua organização de TI é reduzir o custo dessa curiosidade e os esforços necessários a um analista para explorar determinado instinto.

A ideia de Brady se aplica a diferentes organizações e cargos. Um comerciante consegue explorar dados com facilidade para encontrar padrões inesperados na atividade de compra de consumidores? As operações exploram dados para identificar otimizações de desempenho ou diagnosticar problemas em processos operacionais? Agentes financeiros podem explorar dados para inventar novas formas de conduzir seu desempenho ou analisá-los detalhadamente para nortear tomadas de decisão? Líderes de TI conseguem testar suas hipóteses sobre como otimizar gastos na nuvem com rigor e criatividade?

A curiosidade norteia inovação e melhorias. Dados ágeis permitem que funcionários explorem com liberdade suas ideias, instintos, hipóteses e conjecturas na velocidade do pensamento, o que os permite promover novas ideias com suporte de dados.

Para tornar os dados ágeis, uma empresa precisa resolver quais dados deseja e como obtê-los, como preservá-los, como e sob que circunstâncias ela tornará tais dados disponíveis e quais ferramentas e habilidades ela possui para trabalhar com eles.



1

Obter dados



Para usar dados de forma ágil, primeiro precisamos obtê-los. E sabendo que faremos usos desconhecidos desses dados, precisamos coletar *mais* dados do que os que sabemos como usar. Em poucas palavras, isso representa o que “big data” é. Por sorte, com a nuvem, o custo do armazenamento de dados é baixo e está diminuindo. Por isso, nós podemos instrumentar nossos processos empresariais para produzir dados, muitíssimos dados, e torná-los disponíveis para análise. Por exemplo, com frequência aplicativos de Internet das Coisas (IoT) incluem sensores que irrompem fluxos de pontos de dados na nuvem para que a empresa possa analisar de forma imediata ou armazenar para análises futuras. As empresas agora também podem trabalhar com uma gama muito mais ampla de tipos de dados: vídeo, texto e fala, por exemplo. Há muita possibilidade de usar todas essas informações de maneiras novas e interessantes.

A GE Petróleo e Gás, por exemplo, usa em suas tubulações de petróleo um dispositivo do tipo MRI, que a empresa chama de “porco”, para coletar mais de 750 TB de informações que a ajuda a localizar potenciais problemas na infraestrutura de tubulações. A Hudl coletou cerca de 10 PB em vídeo e outros dados que treinadores esportivos podem usar para avaliações junto a atletas. A Peloton reúne dados de seus ciclos de exercícios e os analisa para fornecer insights aos seus clientes. E o Airbnb acumula a cada dia cerca de 50 GB de dados para análise rápida na nuvem, usando o Amazon Elastic MapReduce (EMR), uma ferramenta que permite a análise rápida feita paralelamente de grande volume de dados⁵.

2 | Armazene-os



Uma vez que adquirimos dados, devemos armazená-los para que fiquem disponíveis para análise. Tradicionalmente, armazenamos dados em um formato estruturado com base em nossas expectativas sobre como seriam usados em termos transacionais. Por exemplo, é possível que tenhamos um campo em um banco de dados para a quantidade de pedidos e outro para o preço da unidade. Poderíamos coletar dados para preencher esses campos e arquivá-los, colocando-os nos espaços vagos apropriados no banco de dados, sabendo que sempre poderíamos multiplicá-los por valores para obter um preço total. Ao forçar dados nesse molde, os tornamos úteis para transações, mas podemos perder informações que poderiam ter sido úteis para análise. Esse foi o modelo de banco de dados relacional.

As últimas décadas foram dominadas pelo uso desses bancos de dados relacionais, que são altamente compatíveis com o processamento eficiente de volumes anteriores de dados transacionais, de uma forma que conhecíamos os dados antecipadamente (multiplicar preços de unidades por quantidade de pedidos). Mas ao trabalharmos com dados não transacionais, operando em grandes escalas de transações na internet ou gerenciando dados que não se compartmentam com facilidade em campos de dados pré-definidos, agora há outras alternativas muito melhor projetadas de acordo com os objetivos para uso na nuvem.

Por exemplo, o Amazon Timestream é um banco de dados projetado especificamente para gerenciar dados de séries temporais (como dados produzidos ao longo do tempo por um sensor industrial ou rastreando a atividade de mercado ao longo do tempo); o Amazon Quantum Ledger Database é projetado para o tipo de dados usado na blockchain (dados cujo histórico deve ser verificável usando técnicas como, por exemplo, criptografia); e o Amazon Neptune é projetado para representar conexões e relacionamentos complexos, como redes sociais. As empresas não se limitam mais ao que elas conseguem encaixar em um modelo relacional.

Melhor ainda (em termos de agilidade), dados que serão usados para análises ainda não determinadas podem ser armazenados em um repositório flexível chamado de data lake, onde cada parte deles é armazenada exatamente na forma em que foram recebidos. A potencialidade do data lake reside nas ferramentas que podem ser usadas para analisá-lo: ferramentas que lhe permitem combinar informações heterogêneas, misturando dados estruturados e não estruturados, dados de diferentes silos organizacionais e fontes de grandes quantidades de dados. As ferramentas de hoje podem aplicar algoritmos de machine learning e análises estatísticas e podem trabalhar com linguagem natural como texto, vídeo e fala.

Em outras palavras, o data lake atende a necessidade empresarial de armazenar dados antes de conhecer suas formas de uso. Nós podemos depositar dados no lake a partir de diferentes silos empresariais e analisar todos juntos. Podemos configurar rapidamente uma forma de depositar dados de uma empresa recém-adquirida no lake e, assim, conquistar transparência nas operações a serem estabelecidas com ela, integrando assim seus dados aos nossos. A mágica que torna isso possível é: (1) o baixo custo de armazenamento, (2) a disponibilidade de ferramentas que trabalham com dados heterogêneos pouco estruturados e (3) a disponibilidade de serviços que lhe permitem alimentar dados no data lake com banda larga alta e de forma assíncrona (basta enviar os dados para o data warehouse à medida que você os recebe e eles chegarão o mais rápido possível, sem necessidade de esperar, algo parecido com um e-mail).

3

Torne-os disponíveis



A próxima etapa da viabilização da agilidade de dados é torná-los disponíveis quando e onde forem úteis. Repare que eu não disse quando e onde forem **necessários**. Eu estou falando aqui sobre agilidade e inovação. O modelo comumente usado hoje é de provisionamento de autoatendimento. Quando um analista está curioso, ele pode ativar um conjunto de ferramentas e um subconjunto de dados para realizar análises sem precisar solicitar e esperar alguém fornecer tais dados. Essa consequente liberdade permite que ele adquira um ritmo de pensamento, um fluxo, ao invés de caminhar em uma lógica de interrupções que acaba com a criatividade, ou, você também poderia dizer, que encarece a curiosidade. A nuvem é um recurso importante para tornar isso possível, ela viabiliza o uso de novos ambientes de trabalho, que são descartados adiante quando apropriado. Ela também facilita a aplicação de defesas para proteger a privacidade (mais sobre isso abaixo).

4

Fornecer ferramentas



Uma empresa impulsionada por dados disponibiliza as ferramentas analíticas adequadas aos seus funcionários rápida e facilmente, frequentemente por meio de um modelo de autoprovisionamento, conforme descrito acima. Uma grande variedade de software e serviços está disponível: se você deseja executar consultas tradicionalmente estruturadas em relação a dados, por exemplo, pode configurar um data warehouse com base nos dados presentes no data lake ou pode provisionar uma ferramenta que lhe permita fazer consultas do tipo SQL à moda antiga diretamente em relação ao data lake.

Mas hoje há muitas outras possibilidades. Por exemplo, você pode visualizar seus dados com ferramentas de modelagem e construir cenários que determinam suas consequências. A revolução de análise atual se baseia na inteligência artificial e em machine learning, o que abre novas possibilidades para o que podemos fazer com nossos dados: prever resultados, localizar anomalias, categorizar dados, analisar sentimentos, descobrir padrões, guiar robôs e muito mais.

Por exemplo, a Capital One utiliza machine learning para detectar fraudes enquanto ainda mantém altos níveis de atendimento ao cliente. A T-Mobile utiliza machine learning para melhorar seu atendimento ao cliente de forma que esse recurso faça previsões sobre quais artigos serão mais úteis para clientes e tornando-os rapidamente disponíveis aos agentes de atendimento. A Sky News, durante a cobertura do casamento real britânico, usou o recurso de machine learning da AWS para reconhecer celebridades na multidão e identificá-las para a audiência televisiva. A Formula 1, a Major League Baseball e a National Football League estão usando machine learning para reforçar a experiência do espectador sobre suas atividades esportivas⁶.

Para aplicar o machine learning, você deve treinar um modelo com base em conjuntos de dados anteriores e então aplicá-lo a novos dados, conforme observado. Na AWS, há três abordagens gerais de machine learning: (1) usar um modelo treinado previamente, como o Amazon Rekognition, que já foi treinado para reconhecer objetos e imagens, ou o Amazon Lex, que já foi treinado para compreender intenções expressas em linguagem natural; (2) treinar e aplicar o seu próprio modelo baseado em qualquer um dos algoritmos mais comuns usados para machine learning com o Amazon SageMaker; ou (3) usar os seus próprios algoritmos e abordagens de treinamento, se você tiver funcionários especializados em machine learning, trabalhando diretamente com a infraestrutura da Amazon, que é otimizada para machine learning.

Com ferramentas como essas três, as empresas podem libertar a criatividade de seus funcionários e encontrar novas formas de colocar dados em uso.

5 | Melhorar as habilidades



O próximo elemento importante para extrair valor de seus dados é se certificar de que seus funcionários possuem as habilidades necessárias... além de um senso de curiosidade. É por isso que cientistas de dados estão tão cobiçados hoje em dia. Sim, há uma infinidade de ferramentas disponíveis até para pessoas com poucas habilidades ou experiência com estatística. Mas para realmente fazer um bom uso de dados com rigor, é fundamental contratar pessoas com um bom entendimento de como fazer as inferências corretas a partir de dados.

Um simples exemplo: aqueles que têm menos experiência estatística tendem a confiar demais em médias, até mesmo ao olhar para uma distribuição completa de valores é possível obter insights importantes. Lembro-me de um caso da minha época como diretor de tecnologia da USCIS em que buscávamos reduzir o tempo de processamento de alguns tipos de aplicações. Nós criamos painéis para rastrear a quantidade média de tempo de processamento, mas cada mudança que tentávamos parecia gerar um impacto muito pequeno nas métricas. O que nós não identificamos era que o pequeno número de aplicações que geravam preocupações de segurança ou fraude levava muito mais tempo para ser processado, distorcendo a média. Não tínhamos uma forma de controlar a duração disso. Embora nossas melhorias tenham se aplicado à grande maioria dos casos, por causa da média altamente distorcida, não conseguimos ver o impacto deles. Quando percebemos o problema e começamos a monitorá-lo, por volta do 85º tempo de conclusão perceptível, pudemos identificar o impacto significativo das nossas mudanças na maioria dos casos. Tínhamos os dados, as ferramentas, o acesso... só faltavam as habilidades para delinear inferências corretas.

Decisões orientadas a dados também podem ser fundamentadas de uma forma pobre quando os dados são apresentados (mesmo que de forma não intencional) de uma forma equívoca. Em seu livro *The Visual Display of Quantitative Information*, Edward Tufte mostra como dados podem ser distorcidos ou encobertos pela sua forma de apresentação⁷. Novamente, uma empresa que almeja um uso de dados criterioso precisa garantir que possui as habilidades indicadas para a análise e apresentação, bem como dados em si.

6

Fornecer defesas



Antes de disponibilizarmos dados para novos usos a fim de satisfazer a curiosidade, precisamos aplicar defesas ao redor deles para garantir privacidade e confidencialidade. Empresas impulsionadas por dados exercem a privacidade por projeto, deliberadamente estabelecendo proteções baseadas no planejamento e previsões. Elas adquirem velocidade e flexibilidade nessa estrada se certificando do que precisa de proteção e mobilizando formas automatizadas de garanti-la. Na realidade, a recente Regulamentação Geral de Proteção de Dados (GDPR) exige privacidade por projeto.

A nuvem oferece muitas ferramentas para configurar controles de acesso automático e o faz em níveis granulares que lhe permitem fornecer acesso a funcionários para que acessem de forma precisa os dados necessários. Há formas de rastrear a origem e validade de dados, de criptografá-los ou encobri-los e de restringir o acesso por meio de uma lógica de campo a campo ou registro a registro. Em outras palavras, você pode especificar quais dados de clientes um funcionário pode acessar e quais partes de dados associados aos clientes o funcionário poderá visualizar. O Amazon Macie utiliza machine learning até para identificar quais dados em seu data lake são informações de identificação pessoal (PII) e rastrear como são usadas. Ou você pode escolher gerenciar dados somente em um nível agregado ou com informações disfarçadas ou anônimas. Aí está a flexibilidade; cada empresa impulsionada por dados deve tomar decisões responsáveis sobre a privacidade de acordo com o tipo de dados em uso.

Muitos outros desafios surgem ao utilizar grandes quantidades de dados disponíveis para a empresa. Costuma ser um desafio conectar dados de forma precisa a partir de diferentes sistemas de TI referentes a um único indivíduo, especialmente em países como os EUA, que não possuem um sistema nacional de ID único. Dados podem estar imprecisos não apenas em função de erros cometidos na inserção de dados, como por limitações em sistemas de TI que os coletam. Por exemplo, há sistemas de TI que só permitem um sobrenome e um nome, o que em si já impõe uma imprecisão na presença de pessoas que possuem mais de dois nomes⁸.

Não obstante, o objetivo de empresas impulsionadas por dados é tornar dados disponíveis para nortear tomadas de decisão rigorosas e precisas, assim como a inovação contínua. Isso exige a coleta e o armazenamento de dados para uso posterior de maneira flexível, tornando tais dados e as ferramentas certas disponíveis ao alcance de quem os acessar e sem atrito, garantindo privacidade e confidencialidade por projeto, cultivando as habilidades para fazer inferências válidas e resolvendo os problemas de higiene de dados que podem originar decisões mal informadas. É isso que significa promover agilidade de dados.

Como podemos usar dados para implementar agilidade em nossa empresa?

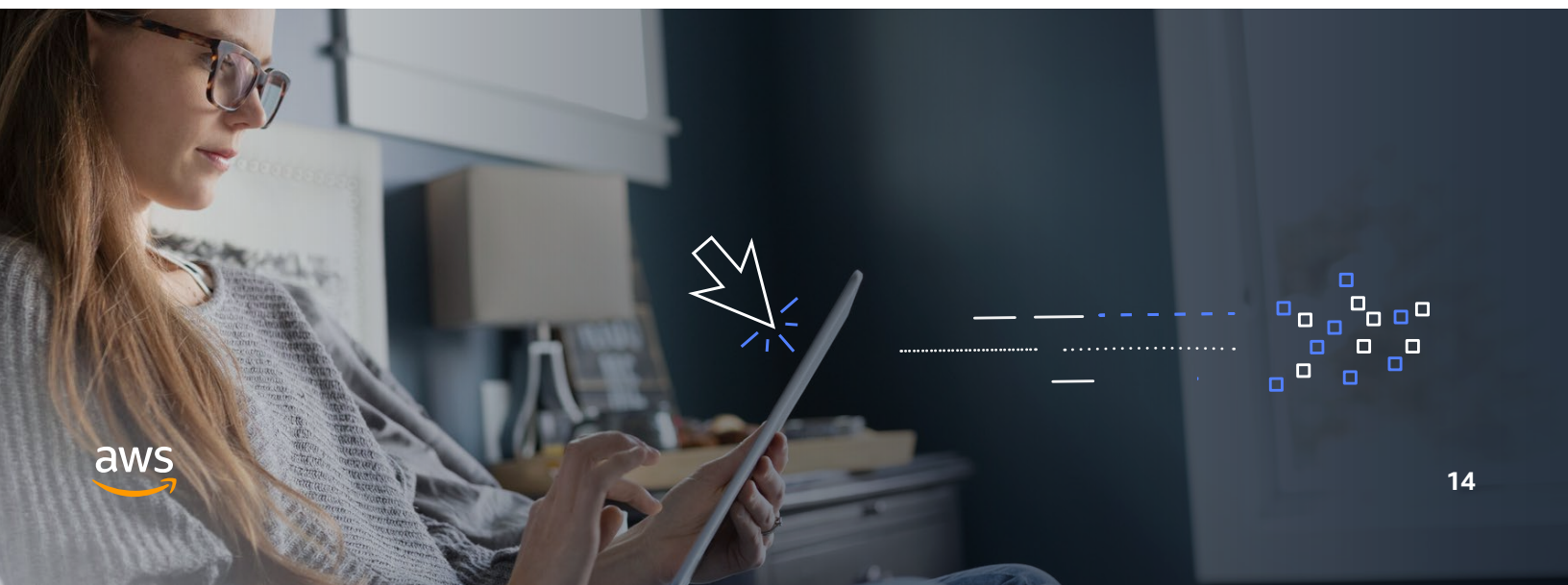
Uma empresa ágil na era digital avança testando uma ideia, buscando feedback e ajustando o curso das coisas repetidamente. Essa abordagem de feedback rápido permite que a empresa inove (com baixo risco, alta velocidade e baixo custo) e reduza seus riscos de investimento testando ideias antes de se comprometer com elas. Isso origina uma boa combinação entre os produtos da empresa e os mercados aos quais ela se interessa em atender e garante que ela solucione o problema correto da forma ideal e na hora certa.

Feedback rápido

Nesse sentido, feedback não significa perguntar aos clientes se desejam um novo recurso ou produto. O mais comum é que empresas impulsionadas por dados usem feedbacks quantitativos, que é o tipo de feedback reunido observando como os clientes *realmente atuam* ou monitorando as mudanças em comportamentos de mercado e outras métricas.

Por exemplo, as empresas costumam melhorar a usabilidade dos seus sites por meio de testes A/B; isto é, experimentando duas variações em uma parte do design (normalmente uma variação da atual, a versão status quo e outra que é uma nova parte do design que elas consideram introdutória). Elas mostram para alguns clientes a versão A e para outros a B. Assim, coletam dados sobre a atividade dos clientes e analisam em função dos resultados que importam em cada caso. Se elas quiserem manter um botão verde ou vermelho para maximizar o número de cliques, alguns usuários verão a versão verde e outros a vermelha e assim a empresa saberá qual cor recebe mais cliques. A Expedia e a Netflix são exemplos de empresas que fazem testes A/B rotineiramente, aproveitando grandes quantidades de dados de um warehouse na nuvem⁹.

⁹ <https://www.youtube.com/watch?v=k8PTetgYzLA>.





A poderosa abordagem de aprender com e fazer ajustes através de feedbacks vai muito além de apenas testes A/B na interface do usuário. Novas ideias de produto, por exemplo, podem ser testadas criando um produto mínimo viável, a menor e mais simples versão do produto que a empresa pode usar para reunir informações sobre como o produto será bem-sucedido ou o que precisa ser modificado para tal. Estratégias de marketing, promoções, alternativas de tecnologia: todos esses fatores podem ser testados por meio de testes e medidas para reduzir inseguranças. A chave para isso é a reunião de dados e a liberação deles para análise.

A técnica do uso de produtos mínimos viáveis e feedback rápido é explicada no livro de Eric Ries, *The Lean Startup*¹⁰. De acordo com Ries, a qualquer momento uma startup mantém duas hipóteses: uma hipótese de valor sobre como o produto oferecido proporcionará valor aos clientes; e uma hipótese de crescimento, que é como a empresa poderá ampliar seus mercados, isto é, conseguir clientes para usar seu produto. O produto mínimo viável é o menor produto que fornecerá à startup informações para confirmar ou refutar tais hipóteses, até que ponto ela poderá promover mudanças e testá-las novamente no mercado.

Esse conjunto de práticas não se aplica apenas às startups ou ao desenvolvimento de novos produtos. Ele se tornou central para a forma em que organizações, incluindo grandes empresas, conquistam a agilidade empresarial mudando o curso de processos com base nas informações inferidas. Se uma empresa estiver pensando em desenvolver um novo sistema de TI para usar com seus próprios funcionários, supostamente ela deve ter uma hipótese sobre como o sistema de TI proporcionará resultados empresariais propostos no caso de negócios em questão. Essa hipótese deve ser testada e as mudanças devem se basear no que os dados evidenciam.

Consequentemente, uma prática ágil exige dados: para aprender e se adaptar, uma empresa precisa coletar dados sobre o impacto de suas novas iniciativas e usá-los para informar tais empreendimentos. Agilidade exige ainda que a empresa perceba mudanças no ambiente empresarial, para que possa responder de forma adequada aos acontecimentos para maximizar seus resultados de negócios. Uma empresa impulsionada por dados não apenas proporciona agilidade aos dados, como também utiliza dados para dar suporte à agilidade.

Mudanças na cultura e nos processos

Nesse sentido, tornar-se impulsionados por dados exige uma forma muito diferente de tomar decisões; para muitas organizações, é uma mudança cultural profunda. No passado, é possível que tenhamos tomado decisões confeccionando planos detalhados, analisando opções junto aos dados disponíveis e escolhendo a opção que, diante dos poucos dados disponíveis, parecia oferecer os melhores resultados. No mundo digital, nos recusamos a aceitar somente os dados disponíveis no momento de criação do plano. Ao contrário, projetamos experimentos para produzir dados adicionais e incorporá-los na tomada de decisões. Resolvemos incertezas gerando novos dados.

Um exemplo é a técnica para governança de TI que desenvolvemos na USCIS. Ao invés de redigir um documento enorme com exigências e entregá-lo aos técnicos da empresa para implementação, nós simplesmente transmitimos um objetivo empresarial. Em um episódio, por exemplo, percebemos que um processador de casos qualificado (um “inspetor de status”) podia processar cerca de 70 casos por dia e o nosso objetivo empresarial era um número muito superior. Em outro episódio, descobrimos que um número de documentos em papel foi perdido em trânsito no deslocamento entre locais de processamento, e nós queríamos eliminar tais perdas.

Para cada um desses objetivos, começamos a criar painéis que indicavam a principal métrica: o número de casos por dia ou o número de arquivos que estavam ausentes. Ao invés de redigir um documento com exigências, criamos uma equipe multifuncional de operadores e técnicos de TI e os cobramos pela melhora da métrica. Nós fornecemos a eles as ferramentas para fazerem mudanças nos sistemas de TI e nos processos empresariais rapidamente e monitoramos juntos tais painéis. Eles fizeram pequenas experiências, mudanças incrementais e monitoraram os resultados todos os dias. Com base no que identificaram a partir dos dados, foram capazes de decidir o que fazer em seguida para maximizar resultados. E o gerenciamento poderia decidir se continuava financiando a iniciativa ou direcionaria os fundos a outros destinos. O resultado foi um processo de governança leve, com riscos reduzidos e orientado a dados que rapidamente proporcionou valor.

Isso nos leva a um outro ponto importante: a responsabilidade é reforçada pela transparência. Tornando dados amplamente disponíveis, nós tornamos o progresso da equipe visível. Como resultado, agentes podem repensar constantemente sua decisão de investimento, investindo mais ou menos, redefinindo objetivos ou interrompendo investimentos de forma definitiva. Os resultados foram o único indicador de sucesso e puderam ser rapidamente encontrados. Mas tais resultados deviam ser suportados por dados.

Padrões de reconhecimento

Outra área em que dados podem proporcionar agilidade é por meio da detecção de mudanças ou reconhecimento de padrões no ambiente. Por exemplo, o machine learning pode ser usado para detectar e responder a anomalias. Nós podemos treinar um modelo de machine learning com dados históricos ou de rotina de forma que este seja usado para o que é considerado normal, em seguida aplicá-lo para encontrar atividades que não sejam normais. Essa técnica pode ser usada, por exemplo, para reconhecer transações fraudulentas ou invasões de rede por hackers. Ou ainda para localizar um equipamento em uma linha de produção industrial que esteja divergindo de seu comportamento normal e possivelmente precise de reparação ou substituição antes que ele falhe.

Quando coletamos grandes quantidades de dados, pode ser que precisemos identificar relações que não sabíamos que existiam. Empresas de mídia social constroem grandes bancos de dados de relacionamentos entre pessoas. A Homeland Security pode identificar que um potencial terrorista que estão investigando já morou no mesmo endereço de outra pessoa já conhecida como terrorista, o que levará os responsáveis a fazer perguntas nesse sentido ao encontrar a pessoa. Pode ser que vários aplicativos de imigração fraudulentos se revelem como tendo sido projetados pelo mesmo advogado de imigração. Aqui, nos movemos para muito além do simples uso de dados para processar transações: agora, podemos encontrar relações importantes e interessantes sobre tais transações. Mas vale reforçar que não sabemos exatamente quais relacionamentos podemos encontrar; agilidade, flexibilidade e curiosidade são as chaves para derivar valor dos dados.

Para citar mais um exemplo de uso de dados para ficar de olho nos eventos, a existência de um ponto de dados pode funcionar como uma confirmação da ocorrência de determinada atividade, como por exemplo quando registros de auditoria são criados automaticamente. Seguindo o registro de atividades, auditores são capazes de validar a conformidade ou investigar atividades inadequadas. A blockchain é frequentemente usada para armazenar dados que confirmam que atividades ocorreram, como uma transferência de dinheiro entre duas partes ou uma aprovação de um contrato entre as partes envolvidas. Ao usar defesas de proteção automatizadas e dados de auditoria para estabelecer a conformidade, as empresas geralmente evitam processos de conformidade mais pesados, que restringem a agilidade.

Claro, existem desafios ao uso de dados para oferecer suporte à agilidade empresarial. Conforme observamos acima, são necessárias habilidades para delinear as inferências adequadas a partir de dados. Os dados nem sempre nos dizem que ação devemos tomar: precisamos interpretá-los e tomar boas decisões. Com frequência, enfrentamos concessões entre falsos positivos e falsos negativos. Por exemplo: se usamos dados para reconhecer transações anormais para identificar potenciais fraudes, corremos o risco de marcar transações demais como anormais e, assim, alarmar clientes; ou de marcar uma quantidade inferior de transações, permitindo que a fraude escape. Quanto maior o conjunto de dados se tornar, maior a probabilidade do crescimento de padrões sem sentido ou de que padrões importantes fiquem ofuscados pelo número absoluto de potenciais conexões. Ruídos se acumulam junto aos sinais.

Para fechar

Uma empresa impulsionada por dados é aquela que coloca dados em uso para melhorar os resultados empresariais, utilizando dados para nortear processos de decisão criteriosos e tornando dados disponíveis para estimular inovação e proporcionar valor aos clientes. Quando dados são armazenados a sete chaves em uma estrutura inflexível, isolados, ou difíceis de acessar, eles se tornam uma barreira para a agilidade, impedindo que a empresa responda a oportunidades ou obtenha produtos do mercado com rapidez. Quando uma empresa *não* conduz seus processos e investimentos utilizando dados, está abrindo mão do importante contato com o mercado ao qual está tentando servir ou recusando feedbacks que poderiam ajudá-la a ser mais bem-sucedida em seus empreendimentos. Uma empresa impulsionada por dados, por outro lado, utiliza dados para conquistar agilidade e utiliza a agilidade para tornar seus dados mais valiosos.

Sobre o autor

Mark Schwartz é estrategista empresarial da Amazon Web Services e autor de *The Art of Business Value, A Seat at the Table: IT Leadership in the Age of Agility e War and Peace and IT: Business Leadership, Technology, and Success in the Digital Age*. Antes de ingressar na AWS, ele foi diretor de TI do Serviço de Cidadania e Imigração (parte do Departamento de Segurança Interna dos EUA), diretor de TI da Intrax e CEO da Auctiva. Ele tem um MBA da Wharton, é formado em Ciências da Computação na Yale e tem um mestrado em filosofia também da Yale.

[Leia mais sobre Mark Schwartz »](#)



Mark Schwartz,
Estrategista
empresarial, AWS