



Três mudanças no ambiente de dados moderno

e as implicações para os líderes de TI

Sumário

O problema é o mesmo, só ficou mais difícil	3
1. Pense em pipelines, e não em buckets	4
2. Use zonas de colocação de dados de acordo com as necessidades.....	6
3. Mudando do protetor de dados para o mentor de dados.....	8
O problema está mais difícil, mas não é impossível de resolver	10

O problema é o mesmo, só ficou mais difícil

Oferecer às organizações dados confiáveis para permitir a tomada de melhores decisões é uma tarefa que, essencialmente, continua a mesma há décadas. Apesar dos grandes avanços tecnológicos e das novas táticas, o departamento de TI que gerencia a infraestrutura de dados atualmente ainda tem a mesma missão básica: movimentar os dados desde o momento da criação e torná-los acessíveis e compreensíveis para os tomadores de decisões sempre que necessário.

No entanto, embora o objetivo continue o mesmo, os obstáculos para se criar e manter uma fonte de informações analíticas em uma empresa se tornaram consideravelmente mais difíceis.

Nos últimos anos, o ponto mais problemático no ambiente de dados moderno talvez tenha sido as novas fontes de dados que geram quantidades sem precedentes de resultados, geralmente com muito pouca (ou nenhuma) estrutura. Desde fluxos de cliques, passando por registros de servidor e fontes de mídia social, até leituras eletrônicas e por sensor, a avalanche de dados provenientes desses canais tem sido literalmente insustentável. Do ponto de vista econômico e de desempenho, os EDWs (data warehouses empresariais) tradicionais são simplesmente incapazes de dar conta dessa onda crescente de dados.

Isso provocou uma total reconsideração das estratégias de coleta e análise de dados e propiciou o surgimento de uma nova geração de soluções de armazenamento de dados voltadas para coleta schema-less (sem esquema), dimensionamento de hardware e colocação da capacidade computacional mais perto dos próprios datastores (e às vezes diretamente neles).

Apesar de ainda serem muito novas se comparadas aos bancos de dados relacionais, essas soluções não relacionais mais recentes cresceram muito nos últimos anos e se desenvolveram rapidamente para dar suporte às maiores e mais complexas empresas do mundo. Embora isso tenha ocorrido em grande parte como forma de complementar as infraestruturas de data warehouse corporativas existentes, acaba por criar um ecossistema de dados mais complexo para os departamentos de TI gerenciarem.

A disponibilidade de dados de aplicativos em nuvem é outra questão que se soma aos desafios que a TI precisa superar na missão constante de manter a integridade do ambiente de dados. Muitas organizações usam aplicativos, como o Google Analytics, o Salesforce, o Netsuite e o Zendesk, entre outros, como componentes essenciais de suas infraestruturas. Os dados que elas geram são fundamentais para os relatórios organizacionais. A capacidade de integrar os dados dessas soluções em nuvem e torná-los acessíveis para a empresa se tornou um requisito básico para os departamentos de TI.

Como o EDW já não funciona como o único destino dos dados, a questão de “quando, onde, como e se” incorporar os dados de aplicativos em nuvem ao ambiente de dados corporativo é um debate constante e acalorado.

Por fim, como as análises de autoatendimento para organizações de todos os portes está se tornando a norma, cada vez mais usuários não técnicos (sem treinamento formal de TI/dados) estão realizando descobertas de dados e gerando relatórios, às vezes até análises preparadas e avançadas. As empresas que adotam esse movimento geralmente observam uma redução significativa (e, às vezes, a eliminação completa) das responsabilidades de produção de análises por parte do departamento de TI. Embora essa mudança seja essencial para o sucesso geral de uma organização que adota uma mentalidade orientada aos dados, ela gera uma nova pressão para as equipes de TI em termos de garantir um acesso mais amplo aos dados. Tudo isso além de garantir que a tecnologia satisfaça as necessidades corporativas e, ao mesmo tempo, atenda às exigências de TI em termos de segurança e governança.

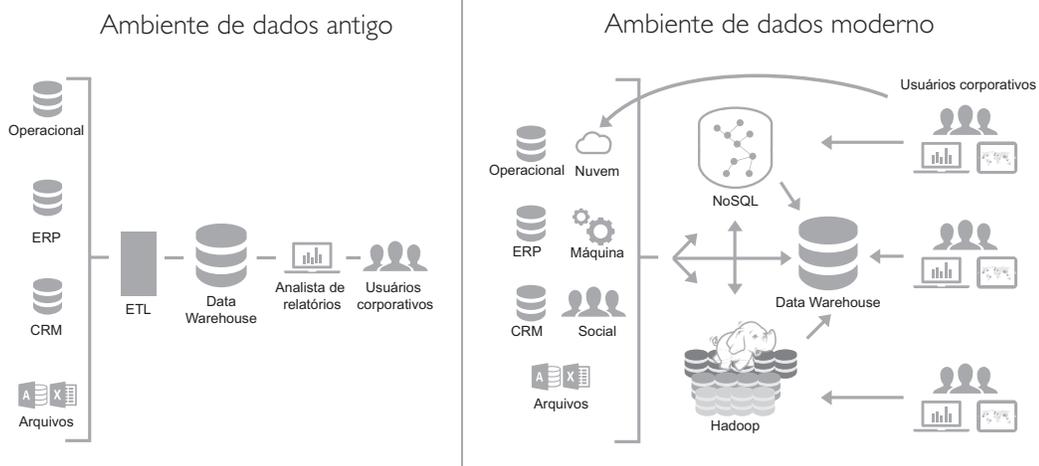
Na tentativa de superar esses novos desafios, muitos departamentos de TI se apressam em adotar novas tecnologias e táticas, mas sem perceberem como esses obstáculos mudaram a forma como as equipes de TI precisam lidar com a tarefa de gerenciar os dados “da criação ao consumo”. Soluções de Big Data, integração de dados em nuvem e análises de autoatendimento são respostas para problemas tecnológicos maiores, mas, a fim de implementá-las de forma eficaz em uma organização, o manual de procedimentos do departamento de TI precisa mudar.

Este documento busca esclarecer as três principais mudanças de pensamento relacionadas ao ambiente de dados moderno que os líderes de TI precisam entender para permitir a tomada de decisões orientadas por dados em suas organizações.

1. Pense em pipelines, e não em buckets

O data warehouse empresarial não acabou. Ele apenas tem mais amigos. De clusters do Hadoop a bancos de dados NoSQL, o EDW relacional já não é o único local oficialmente aprovado para armazenar os dados.

Isso significa que o processo de movimento dos dados não precisa mais ser centralizado em torno de um único local. Na realidade, hoje, uma plataforma de dados moderna é algo mais parecido com o tecido conjuntivo nervoso que liga os hemisférios do “cérebro de dados” de uma organização. Some isso às soluções em nuvem atualmente onipresentes que permitem que infraestruturas e serviços implementem projetos de pipeline/ETL em questão de horas e temos uma receita para movimentar os dados por uma organização como nunca havia sido possível.



Infelizmente, muitas equipes de TI perdem essa oportunidade, pois grande parte do manual de procedimentos de gerenciamento de dados de muitas organizações foi escrita de acordo com a mentalidade ultrapassada do “bucket único de dados”.

O segredo para mudar de uma mentalidade de “bucket” para uma de “pipeline” é aceitar que nem todas as questões relacionadas a dados de uma organização podem ser respondidas por uma única fonte de dados. Um EDW talvez não seja capaz de oferecer a segmentação de dados necessária para filtrar e-mails acumulados nos últimos cinco anos diretamente em registros do servidor de mensagens. Além disso, uma implantação do Hadoop talvez não seja capaz de oferecer respostas a consultas de dados de vendas na velocidade necessária para análises em tempo real. No mundo real, muitos problemas empresariais exigem ao mesmo tempo segmentação de dados e velocidade de consulta de uma ou várias fontes – porém, em períodos diferentes e em sucessões variáveis para que seja possível concluir um projeto de dados.

Levando em conta esses requisitos, a pergunta para a TI já não é “em qual bucket esses dados devem estar?” Em vez disso, a pergunta é “qual estágio dos dados é necessário?” e “como permitir que os usuários transitem facilmente entre os dois?”.

No final das contas, estamos falando sobre pipelines, pois a localização real de determinada fonte de dados pode mudar ao longo dos diversos estágios do processo. Além disso, a mentalidade de pipeline vai muito além do âmbito da integração de dados tradicional e diz respeito ao próprio fluxo de trabalho organizacional.

Como os usuários corporativos estão encontrando suas respostas associadas a dados? A maioria das equipes de TI sabe que existem repositórios de dados individuais e não aprovados mantidos em planilhas por toda a organização. Porém, os departamentos de TI têm se sentido impotentes para conter a propagação de processos de gerenciamento de dados fragmentados, não controlados e em arquivos locais. Muitas vezes, os funcionários das empresas sentem que ignorar seus respectivos sistemas de dados corporativos é a única solução prática – e fazem exatamente isso.

A mudança para a mentalidade de pipeline só ocorrerá quando as equipes de TI entenderem por completo como os usuários corporativos encontram, filtram, analisam e apresentam os dados – e se tornarem capacitadoras ao longo do processo (mas sem deixarem de incentivar a governança). Uma mudança fundamental e crítica ocorre quando as equipes de tecnologia incorporam a ideia de que o movimento dos dados, tanto no sentido tradicional de TI em lotes quanto na natureza de consultas sob demanda dos usuários corporativos, é mais importante do que o esconderijo de dados fortificado.

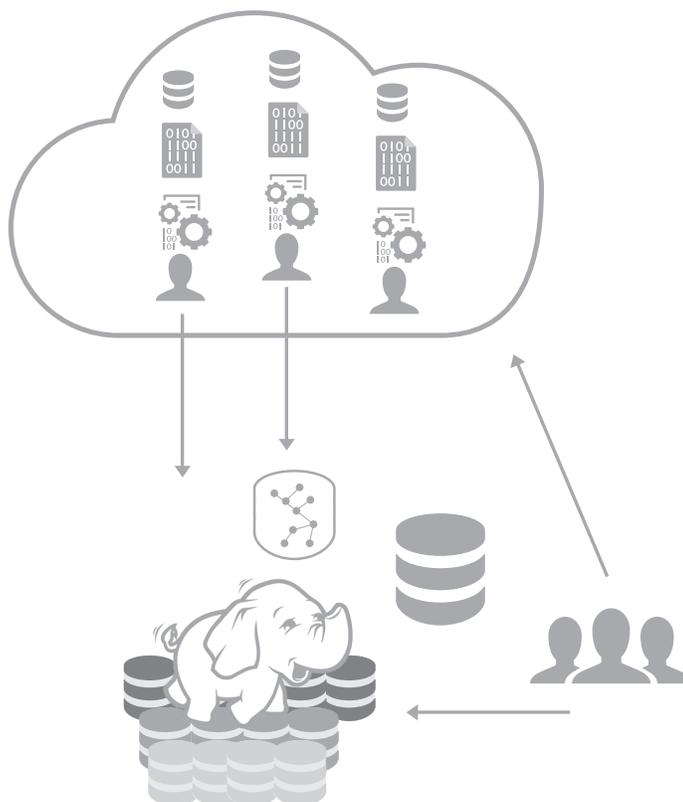


De uma perspectiva global, é cada vez mais importante ter acesso a todas essas informações e poder entendê-las. Agora, com o que criamos usando nosso data warehouse lógico e o Tableau, as pessoas podem colaborar, e todas essas informações estão acessíveis.

— TIM NALL, CIO, BROWN-FORMAN

2. Use zonas de colocação de dados de acordo com as necessidades

Os dados de aplicativos em nuvem nem sempre devem ser colocados diretamente no data warehouse empresarial. A decisão deve sempre condizer com as necessidades da organização. Na realidade, a decisão deve remeter ao que é necessário para tornar os dados úteis para a organização.



Um engano comum entre as equipes de TI, mesmo naquelas que rejeitaram a abordagem do “único bucket de dados”, é pensar que os dados de aplicativos Web devem ser automaticamente integrados ao EDW de forma direta. À primeira vista, isso faz sentido, considerando que muitos aplicativos em nuvem corporativos de sucesso são partes integrantes da infraestrutura de vendas, marketing e suporte. Esses dados são altamente acionáveis, amplamente acessados e geralmente bem estruturados. O ROI (retorno sobre o investimento) de colocar esses ativos no EDW, cujo preço por byte é altíssimo, parece sólido.

No entanto, as perguntas fundamentais que as equipes de TI precisam fazer são:

- Os dados estão prontos para análise diretamente na nuvem? Em outras palavras, eles estão filtrados e devemos nos dar ao trabalho de transferi-los?
- O valor real desses dados na nuvem só é obtido depois que eles são combinados com outros dados?

Dependendo das respostas a essas perguntas, o local onde a equipe de TI armazena seus ativos de dados em nuvem (isso se optarem por trazê-los para o ambiente local) pode variar dependendo da organização e do aplicativo.

Qualquer equipe de TI responsável por uma implantação de tamanho considerável do Salesforce, com poderá lhe dizer imediatamente que permitir registros de oportunidades duplicados no EDW é uma receita para o fracasso. Desde pagamentos de comissão imprecisos a números de receita extremamente superestimados, esses problemas são apenas “a ponta do iceberg” quando se trata de lidar com dados de vendas filtrados de forma inadequada.



Ao colocar os dados nas mãos dos nossos usuários finais, eles foram capazes de reunir rapidamente seus próprios relatórios de nível básico. Eles eram as pessoas mais próximas dos produtos e das alterações no aplicativo e puderam identificar rapidamente em que ponto algo poderia precisar de ajustes.

— SHARON GRAVES, DIVULGADORA DE DADOS EMPRESARIAIS, GODADDY

Da mesma forma, dados de sites e marketing de fontes em nuvem, como Google Analytics, Eloqua e Marketo, podem prejudicar terrivelmente a capacidade de uma organização de rastrear o fluxo de oportunidades e de calcular os custos de aquisição para os clientes se a precisão dos registros não for garantida.

Esse problema de precisão se torna muito pior quando é necessário mesclar as fontes. Esse geralmente é o caso quando uma organização precisa traçar um quadro completo de seu funil de clientes – desde o primeiro acesso a um site à decisão de vendas. Embora quase todas as equipes de TI conheçam a importância de dados filtrados, seja da nuvem ou de outras fontes, elas geralmente não levam em consideração onde eles serão colocados primeiro após serem extraídos de um aplicativo Web.

Um departamento de TI com um ambiente de dados de vários buckets/com foco no pipeline avaliará a integração dos dados em nuvem com base no valor dos ativos de dados para a empresa no momento que eles forem transferidos para o ambiente local. Quando uma fonte de dados em nuvem está relativamente filtrada e agrega muito valor à empresa logo após a implementação, o melhor conselho para as equipes de TI é colocá-la onde ela possa ser acessada com mais rapidez (EDW relacional).

No entanto, ao lidar com dados de aplicativos Web que exigem processamento intensivo e/ou complexo antes de serem amplamente utilizados com confiança pela empresa, as equipes de TI podem aproveitar ambientes de alta computação e baixo custo por byte, como um ambiente do Hadoop. Com essa segunda abordagem, os recursos podem ser mais bem aproveitados com a filtragem e a transformação de dados sem afetar a velocidade do EDW. A partir daí, o departamento de TI que trabalha com a empresa pode decidir se os dados filtrados devem ser transferidos para um EDW e/ou acessados diretamente no ambiente maior.

Por fim, é possível que não seja necessário transferir os dados do aplicativo em nuvem. É cada vez mais comum que os aplicativos Web ofereçam pontos de acesso rápido para seus repositórios de back-end, ou seja, os usuários corporativos podem usar ferramentas de relatório e análise de autoatendimento para realizar suas próprias investigações nos dados em tempo real. Para os departamentos de TI que ainda estão interessados em ter uma camada intermediária de autorização e controle nesses casos, algumas das soluções de análises de autoatendimento voltadas para empresas também permitem um cenário de conexão via proxy. Isso pode atender a todas as necessidades do departamento de TI, desde acesso básico do usuário, até lógicas de negócios altamente complexas. Considerando esse fato, a primeira pergunta sobre a integração de dados de aplicativos em nuvem não é “onde?”, mas sim “será que devo?”. Há uma necessidade convincente de transferir os dados (filtragem, valor adicional etc.), ou eles podem ser deixados no local e possivelmente integrados no nível do usuário corporativo?



Escolhi o Tableau por sua solidez e seu roteiro claro para o futuro. E esse roteiro está alinhado com o caminho que eu quero seguir... ele me ajuda a manter o ritmo que desejo e preciso.

— STEVEN JOHN, CIO, AMERIPRIDE

3. Mudando do protetor de dados para o mentor de dados

O movimento rumo à análise de autoatendimento nas empresas deve ser visto como um acordo com os departamentos de TI, a ser selado com um aperto de mãos, e não uma ameaça de invasão hostil. As equipes de TI que estenderem a mão primeiro só têm a ganhar.

Da mesma maneira que o ambiente de dados moderno se expandiu para incorporar mais do que apenas datastores relacionais, o ambiente de análise também cresceu para incluir ferramentas que agregam um imenso valor para os usuários corporativos típicos. Isso causou a ampla eliminação do processo tradicional em que todas as análises passavam por um pequeno grupo de pessoas altamente capacitadas.

Uma parceria que funciona

Função do departamento de TI

- Segurança
- Arquitetura de dados
- Dimensionamento
- Treinamento
- Centro de operações
- Intranet de habilitação

Função da empresa

- Trabalho analítico criativo
- Condução da aquisição de dados
- Compartilhamento de conhecimentos
- Busca de ajuda quando necessário
- Divulgação
- Incentivo à ação

Habilitação



Execução

Como resultado, equipes de TI visionárias estão se transformando em líderes de inovação em dados e análises em suas organizações e redefinindo a maneira como trabalham com as equipes corporativas para resolver os desafios associados a dados. Essa mudança de mentalidade do departamento de TI de “protetor” para “mentor” de dados talvez seja a transição mais marcante no ambiente de dados moderno. Esse é um importante fator no sucesso ou fracasso da iniciativa de uma organização em se tornar realmente orientada pelos dados.

A redefinição desse relacionamento entre o departamento de TI e a empresa implica ajustar a atitude padrão de governança de dados de “não” para “sim”. Antes, o acesso de um usuário corporativo aos dados era automaticamente negado, a menos que ele tivesse aprovação explícita. Esse novo modelo pressupõe que qualquer usuário corporativo de qualquer parte da empresa possa acessar qualquer ativo de dados, desde que não viole as normas de conformidade.

Os efeitos dessa mudança de atitude são enormes. O impacto pode ser visto na maneira como os usuários corporativos começam a buscar novos dados para solucionar novos problemas. Sua disposição para isso tem uma relação direta com sua percepção de quanto esforço será necessário para acessar novos dados. Quando equipados com uma solução de análise de autoatendimento fácil de usar e incentivadas pelo conhecimento de que seu departamento de TI apoia o amplo acesso aos dados, os funcionários da empresa podem se tornar agentes de mudança. E esses agentes de mudança são indicadores da entrada bem-sucedida nos primeiros estágios do processo de se tornar uma empresa orientada pelos dados.

No entanto, o simples fato de liberar o acesso aos dados e distribuir novos softwares não resolve o problema maior de realmente ajudar a empresa a tomar melhores decisões com os dados. Os departamentos de TI que colaboraram com êxito para migrar suas empresas para mecanismos de decisões orientadas por dados fizeram isso ao realmente incorporarem o papel de mentores de dados. Muitas vezes, os membros da equipe de TI que antes controlavam o processo de análise se tornam guias para o restante da empresa. Eles têm as habilidades para treinar outras pessoas em funções essenciais, como análise adequada de dados e criação de relatórios.

Educar os usuários corporativos dessa forma amplia de forma eficaz e ampla a capacidade de criação de relatórios analíticos de toda a organização. Isso também ajuda as organizações a obter retornos maiores e mais rápidos sobre seus investimentos em soluções de análises de autoatendimento. No entanto, o mais importante é que o uso dos dados para entender melhor a empresa como um todo se torna parte do DNA operacional da organização.

O problema está mais difícil, mas não é impossível de resolver

Desde a necessidade de lidar com soluções de armazenamento e processamento de dados totalmente novas à compreensão de quando, como e se os dados em nuvem devem entrar no ecossistema, as responsabilidades das equipes de TI de gerenciar o fluxo de dados organizacional estão rapidamente se tornando mais difíceis. Some isso ao movimento de análise de autoatendimento e essas mudanças juntas se tornam algo maior – uma profunda transformação em como as empresas lidam com os dados.

As organizações precisam ser guiadas pelo departamento de TI ao longo dessa mudança. Grande parte da incerteza que as equipes de TI sentem se deve muito à noção de que o problema já não tem a ver com adaptar novas tecnologias aos processos corporativos existentes. Big Data, tecnologias em nuvem e a habilitação de um modelo de análise de autoatendimento representam mudanças fundamentais em como as equipes de TI precisam abordar o ambiente de dados moderno.

Para as organizações que já começaram a reavaliar sua mentalidade sobre dados e análises a fim de adaptar-se a essas mudanças, os obstáculos são concretos – mas as descobertas também são. A mudança é um desafio tanto para os usuários corporativos como para o departamento de TI. Durante a redefinição do relacionamento entre as duas partes, elementos de ambos os grupos demonstrarão resistência. Porém, também haverá aqueles que aceitarão esse processo com entusiasmo.

À medida que os muros e os processos rígidos que rodeiam os dados organizacionais passam por mudanças, nos quatro cantos da empresa, as equipes de TI encontrarão defensores inesperados da análise ansiosos para fazer parte da vanguarda de uma nova empresa orientada pelos dados. Os melhores líderes de TI juntarão esses novos defensores aos seus experientes profissionais de dados, transformando-os também em mentores. Os defensores aprenderão a lidar com o ecossistema de dados em evolução e o departamento de TI identificará os problemas corporativos que precisam ser resolvidos. Essas primeiras etapas estabelecem as bases para uma cultura de análise e descoberta de dados em uma empresa na qual as decisões são fundamentadas pela confiança e pela precisão do ambiente de dados moderno.

Sobre o Tableau

O Tableau ajuda as pessoas a ver e entender seus dados, independentemente do volume, do canal de origem ou do banco de dados em que estão armazenados. Conecte, combine e visualize rapidamente seus dados com uma experiência integrada, do PC ao iPad. Crie e publique painéis de marketing com atualizações de dados automáticas e compartilhe informações em tempo real com colegas, equipes, líderes executivos, parceiros ou clientes, sem precisar ter conhecimentos de programação.

Recursos adicionais

Obtenha uma avaliação gratuita

+ [Tableau Online](#)

+ [Tableau Desktop](#)

Whitepapers relacionados

[Guia de avaliação de BI e análises modernas](#)

[Análise de autoatendimento governada e escalonável](#)

[Redefinindo a função da TI no mundo da abordagem moderna de BI](#)

[Tableau para a empresa: uma visão geral de TI](#)

