



Componentes integrantes de uma plataforma de análise moderna

Dan Kogan e Jen Underwood

O que é “análise moderna” e por que precisamos dela?

Vivemos em um momento empolgante de rápida inovação, maior concorrência global e oportunidades inéditas de revolucionar e reinventar. A quarta revolução industrial é possibilitada por fatores como o crescimento vertiginoso da tecnologia digital, a conectividade com a Internet em qualquer lugar e a proliferação de dispositivos inteligentes. Tudo isso cria uma enxurrada de dados que trazem uma vantagem competitiva para aqueles que podem entendê-los com rapidez e precisão.

O potencial da tecnologia moderna não deve ser subestimado. A linha definida entre os mundos físico e virtual está desaparecendo, o que facilita a entrada no mercado e cria modelos de negócios totalmente novos. Todos os setores estão prontos para essa revolução, até mesmo aqueles que, há uma década, foram os revolucionários. Considere o impacto do crowdfunding sobre os empréstimos para pequenas empresas, do varejo on-line sobre o grande mercado varejista ou dos serviços de mídia por assinatura sobre a televisão a cabo tradicional. Em toda parte, essa transformação está se acelerando: processos automatizados, bots de bate-papo e inteligência artificial são apenas algumas das novas tecnologias que estão abrindo caminho para um futuro fascinante, porém incerto.



Para saber mais sobre o papel mutável do termo mais amplo e popular do BI, leia nosso whitepaper **“Definição de análise”**, que aborda os principais termos utilizados no contexto da análise atualmente.

No entanto, há um fio comum que permeia todas essas tendências: volumes imensos de dados.

Se os dados são a matéria-prima, a análise é o processo de refiná-los para produzir informações úteis e, assim, proporcionar vantagem competitiva à sua empresa. Hoje os dados são mais indispensáveis do que nunca. À medida que as empresas evoluem para acompanhar o ritmo dos mercados em constante mudança, elas precisam recorrer a volumes crescentes de dados precisos e atuais para tomar decisões rápidas e inteligentes. Porém, a tecnologia de análise sempre foi lenta e difícil de adaptar ao crescimento e aos formatos mutáveis dos dados. A cada semana, parece haver um tipo diferente de banco de dados, com dados gerados por novos dispositivos e criados com tecnologias inéditas. A análise simplesmente não conseguiu acompanhar essa evolução.



Para saber mais sobre como o trabalho conjunto da equipe de TI e da empresa é fundamental para possibilitar a abordagem moderna de análise empresarial, leia nosso whitepaper: **Como desenvolver uma cultura de análise**.

Na atual era digital, para promover uma cultura impulsionada por dados capaz de solucionar os desafios comerciais modernos com precisão e velocidade, as organizações precisam investir não só em novas tecnologias, mas também em novas maneiras de distribuir essas informações, incluindo as pessoas necessárias para conduzir essa missão. Essa transição cultural simultânea representa uma mudança fundamental na relação entre a equipe de TI e a empresa, duas parceiras que trabalham para coletar e minerar os dados, mas também para lapidá-los e disponibilizar as informações certas conforme necessário. É somente quando a equipe de TI e a empresa trabalham juntas que as organizações podem tornar realidade o sonho de uma cultura de análise de autoatendimento.



Quando os departamentos de TI conduzem a transição para a análise de autoatendimento, eles podem garantir a escalabilidade da governança e da segurança. Além disso, ao capacitar a empresa a ser ágil e orientada pelos dados, a TI se torna uma parceira de confiança da empresa.

- COLIN REES, CIO, DOMINO'S PIZZA GROUP LIMITED

Este whitepaper examinará os componentes integrantes de uma plataforma de análise moderna que a empresa e a equipe de TI podem usar juntas para disponibilizar os dados, agregar valor e tomar decisões em prol de toda a empresa. Isso inclui as novas e extraordinárias ferramentas disponíveis atualmente, bem como as ferramentas tradicionais que, há décadas, são as bases do business intelligence. Mostraremos como cada um desses componentes integrantes se encaixa no processo mais amplo de transformar os dados em informações, sejam as ferramentas necessárias para coletar os dados e usá-los para criar relatórios ou as novas tecnologias para compartilhar informações interativas. Também abordaremos como o Tableau pode ser a base para uma plataforma de análise moderna e, ao mesmo tempo, um catalisador para a criação de uma nova cultura de análise impulsionada por dados.

Sumário

O que é análise moderna e por que precisamos dela?	2
1. Os três grandes desafios atuais em torno dos dados	4
Os dados estão em toda parte	4
Todos precisam dos dados	5
Os dados estão sempre mudando	5
2. Os componentes integrantes de uma plataforma de análise moderna	6
Habilitação pela TI	7
Criação e consumo	9
3. Testando todos os componentes em conjunto	12
4. Apêndice.....	13
Processamento em fluxo contínuo	13
Orquestração de hubs de integração.....	13
Mais informações sobre dados desestruturados, NoSQL e lagos de dados	14
Dados como serviço.....	15
Data warehouse lógico.....	15
Aprendizado de máquina.....	17
Linguagem natural	17
Descoberta de dados recomendada.....	17
Pesquisa	17
Alertas	18
Histórias	18
Sobre o Tableau	19

1. Os três grandes desafios atuais em torno dos dados

Os dados estão em toda parte, e grandes volumes de dados são gerados a cada segundo. Por exemplo, uma escova de dentes agora é um dispositivo inteligente capaz de registrar quando você escova seus dentes, o tempo de escovação e o status das peças internas. Ela também pode enviar todas essas informações para o seu dentista. Uma tarefa banal se transformou em mil pontos de dados. Multiplique esse exemplo por milhões de dispositivos e você terá um pequeno mercado individual que cria montanhas de dados que simplesmente não existiam há alguns anos. Considere registros de eventos, APIs, redes sociais, acompanhamento de sites e uma série de outras tecnologias on-line e há uma explosão de dados a cada vez que você pisca.

Esse ecossistema moderno apresenta três desafios comerciais:

1. Os dados estão em toda parte
2. Todos precisam dos dados
3. Os dados estão sempre mudando

Os dados estão em toda parte



Figura 1 O Tableau se conecta aos dados em qualquer lugar

Historicamente, as organizações sempre mantiveram os dados na infraestrutura local. Elas se empenhavam para controlar todos os dados que eram criados e armazenados em um data warehouse aparentemente bem definido. Se havia algum dado que não era coletado, ele provavelmente era irrelevante e o esforço não valia a pena.



Para saber mais sobre as tendências atuais de migração de dados, leia nosso [Resumo sobre os dados na nuvem](#).

Essa mentalidade pode acabar com um negócio na época atual, em que sites, dispositivos móveis e aplicativos na nuvem geram dados fora dos limites da organização. Considere o Google Analytics, o Splunk, o ServiceNow e o Salesforce, para citar apenas alguns exemplos. Essa tendência só deve acelerar, e cada vez mais dados úteis serão gerados na nuvem por provedores terceirizados. Para acentuar essa inclinação à nuvem, algumas organizações também estão migrando sua própria infraestrutura local para a nuvem.

Todos precisam dos dados

Nas últimas duas décadas, vimos o mercado passar por uma transição para os negócios digitais e a nuvem. Esse é um fato importante na revolução da análise moderna. A outra transição envolve a criação de uma cultura impulsionada por dados com o business intelligence de autoatendimento. Uma cultura de análise já existe na maioria das empresas mais inovadoras da atualidade, nas quais um usuário corporativo que tem uma pergunta pode descobrir a resposta por conta própria. Isso capacita as organizações a transformar os dados, que são o ouro digital, em informações com imensa rapidez. Para criar uma cultura de análise, as organizações devem unir pessoas e dados, que muitas vezes são considerados os bens mais valiosos de uma empresa, dando a todos acesso aos dados certos e estimulando-os a explorar e colaborar.

Com uma abordagem moderna de análise, a equipe de TI e a empresa trabalham juntas. A equipe de TI disponibiliza um ambiente centralizado no qual os usuários corporativos podem encontrar dados e conteúdos confiáveis, permitindo que todos utilizem esse ambiente com segurança, façam perguntas, experimentem e tomem decisões na velocidade do pensamento. Trata-se de uma metodologia "de baixo para cima" em que os especialistas técnicos criam metadados, regras de negócios e modelos de relatórios que proporcionam agilidade ininterrupta e aceleram a melhoria contínua.



Quando implementamos o Tableau, tínhamos apenas a geração de painéis e relatórios em mente. Não imaginávamos que ele poderia mudar radicalmente o DNA da organização. Não se trata apenas de uma solução ou de um tipo de tecnologia, e sim de como o comportamento em relação aos dados mudou.

- ASHISH BRAGANZA, DIRETOR DE BUSINESS INTELLIGENCE GLOBAL

Os dados estão sempre mudando

Como todos sabemos, a única verdadeira constante é a mudança. Uma plataforma de análise moderna prioriza a flexibilidade: a capacidade de transferir os dados entre plataformas, ajustar a infraestrutura conforme necessário, aproveitar novos tipos de dados e incorporar novos casos de uso. Além disso, parece que uma nova tecnologia de análise de dados é lançada no mercado todos os dias: tecnologias como o aprendizado de máquina, assistentes de voz e consultas em linguagem natural. Algumas delas podem parecer mais ficção do que realidade, pelo menos por enquanto, mas novos métodos e técnicas se desenvolverão e provarão seu valor para sua empresa e seus clientes.

Em um mundo onde os dados evoluem rapidamente, a flexibilidade é fundamental para acompanhar suas necessidades de infraestrutura em constante expansão e também para as novas tecnologias. A flexibilidade é vital para conquistar e conservar sua vantagem competitiva. Ao considerar uma arquitetura de análise compatível com o futuro, evite a dependência de fornecedores específicos com arquiteturas comerciais, pois isso pode comprometer sua capacidade de ser ágil no futuro.

2. Os componentes integrantes de uma plataforma de análise moderna

Esses três desafios que as empresas enfrentam hoje são menos insuperáveis do que parecem. Se os dados são o fio comum do mundo dos negócios em constante evolução, uma plataforma de análise moderna é indispensável para explorar todo o seu potencial. No entanto, a plataforma de análise moderna não é um elemento individual da infraestrutura: ela é formada por vários componentes integrantes independentes. Alguns deles são as bases tradicionais do business intelligence adaptadas à era moderna (por exemplo, os data warehouses). Outros, são conceitos totalmente novos que revolucionaram a forma como as empresas abordam a análise de dados (por exemplo, a análise visual). Juntos, esses dois tipos de componentes integrantes formam uma plataforma de análise que pode ajudar qualquer organização a enfrentar os desafios de negócios atuais.

Uma plataforma de análise moderna pode ser dividida em duas partes distintas:

- **Habilitação pela TI**, que inclui a coleta, a seleção e a preparação dos dados
- **Criação e consumo**, que inclui analisar os dados e divulgar as informações descobertas para as partes interessadas.

Essas duas partes sempre foram combinadas em um único processo que era designado à equipe de TI e a ninguém mais. Agora vemos a primeira parte (a criação e o processamento das fontes de dados) como aquela habilitada pela TI. A segunda parte (análise e distribuição) continua sendo viabilizada pela TI, mas é conduzida pelos próprios usuários corporativos.

A combinação dessas duas partes representa uma verdadeira parceria entre a empresa e a equipe de TI. Essa é a forma moderna de administrar qualquer organização que deseja ser capaz de tomar decisões impulsionadas por dados com rapidez. Às vezes, essa abordagem é chamada de "BI bimodal", preservando o melhor do BI e dos relatórios operacionais tradicionais, mas, ao mesmo tempo, adotando o mecanismo de autoatendimento da análise moderna.

Nessa relação entre a equipe de TI e a empresa, a TI desenvolve a arquitetura de dados e garante a segurança dos dados e os controles de acesso adequados. Os especialistas técnicos da empresa então criam os recursos analíticos de que precisam, quando precisam deles. Resultado: a equipe de TI capacita todos a responderem suas próprias perguntas importantes e os usuários corporativos podem responder a uma pergunta no momento em que ela surge. Isso cria uma organização ágil preparada para vencer os desafios do mundo moderno dos negócios.

Faremos uma análise geral dos vários componentes integrantes que formam cada parte da plataforma de análise moderna, algumas das principais tendências em cada um deles e conceitos importantes que devem ser lembrados. Para conferir uma descrição detalhada dos componentes específicos, incluindo os fornecedores líderes de mercado que os oferecem, e para saber se eles se aplicam ao seu caso, consulte a seção correspondente a cada componente integrante no [Apêndice](#).



Figura 2 Componentes integrantes básicos

Habilitação pela TI

Ao contrário do business intelligence tradicional conduzido pela equipe de TI, os departamentos de TI mais eficazes da atualidade se dedicam a habilitar fontes de dados analíticas orquestrando, organizando e unificando os dados para que os usuários e especialistas possam criar e consumir. Essa função é óbvia, mas deve ser enfatizada. A tarefa de coletar dados, gerenciar suas fontes e processá-los para serem usados pelas pessoas sempre foi fundamental para o business intelligence, e continua sendo uma parte essencial da plataforma de análise moderna. Para que haja informações a serem descobertas, antes é necessário refinar toda essa matéria-prima.

O diferencial da plataforma de análise moderna é a parceria entre a empresa e a equipe de TI. Quando os usuários corporativos dispõem de ferramentas para analisar os dados por conta própria, eles têm a liberdade de responder as perguntas à medida que elas surgem, sabendo que podem confiar nos dados disponíveis. Isso possibilita a criação de relatórios e painéis ágeis e precisos. Ao mesmo tempo, sem a tarefa de atender às solicitações de alteração e criação de painéis, a equipe de TI finalmente pode priorizar os dados: garantindo a governança e a segurança dos dados, assegurando sua precisão e estabelecendo os canais mais eficientes para coletar, processar e armazenar os dados.

Este é o melhor momento de priorizar os dados. Sua empresa, independentemente do porte, já coleta dados e provavelmente já analisa uma pequena parte deles: o resto são **dados obscuros**. Há bilhões de lugares para coletar dados e cada vez mais ferramentas são lançadas no mercado para ajudar na coleta do maior volume de dados possível. Hoje, há uma infinidade de tecnologias sendo utilizadas para tratar vários aspectos dos dados, como alto volume, localização e uma variedade de tipos de fontes de dados. Cada organização é única e você deve identificar e priorizar os componentes que são mais relevantes para sua empresa hoje e no futuro.

Veja a seguir alguns pontos importantes.

Para conferir mais detalhes sobre tecnologias específicas, como dados em fluxo constante e dados como serviço, incluindo opções de fornecedores específicos, consulte o [Apêndice](#) a seguir.

Bancos de dados e data warehouses

Os bancos de dados e data warehouses foram as bases do business intelligence durante décadas. Alguns deles ainda têm seu lugar na arquitetura de análise moderna, enquanto outros estão começando a perder sua relevância.

Alguns dos bancos de dados mais antigos são conhecidos como OLAP (processamento de análises on-line). Eles foram desenvolvidos como uma alternativa a tecnologias de banco de dados lentas, e usavam agregações e armazenamento em cache para acelerar o tempo de resposta de consultas previsíveis. Porém, como as perguntas das empresas se tornaram mais complexas e difíceis de prever, o padrão OLAP está ultrapassado, pois muitas vezes requer a criação de agregações totalmente novas. Ele também perde relevância diante de tecnologias de banco de dados mais sofisticadas.

Os bancos de dados modernos aproveitam os avanços computacionais, como a tecnologia MPP (processamento paralelo maciço) e o processamento na memória. Com isso, os bancos de dados podem oferecer um desempenho extremamente rápido com escalabilidade linear, além de otimizar o armazenamento de dados e o uso da memória física. Às vezes, podem até incluir funções computacionais e de ciência de dados integradas.

Além disso, o advento da computação em nuvem trouxe grandes avanços para as tecnologias de banco de dados que as versões locais simplesmente não podem superar. Isso inclui a possibilidade de começar sem adquirir hardware, de escalonar a capacidade de forma elástica à medida que as necessidades comerciais mudam e de gerenciar a infraestrutura sem montar uma equipe.

Sempre haverá lugar para bancos de dados e data warehouses na arquitetura de análise moderna, e esses componentes continuam desempenhando um papel fundamental no fornecimento de dados dimensionais governados, precisos e adequados para a criação de relatórios de autoatendimento em toda a empresa. Até mesmo empresas que adotam outras tecnologias (como o Hadoop e lagos de dados) geralmente preservam os bancos de dados como parte de seu conjunto de fontes de dados.

NoSQL, dados desestruturados e lagos de dados

Os bancos de dados e data warehouses são particularmente úteis para a análise quando os dados têm fontes de dados e formatos previsíveis. Obviamente, nem todos os dados são previsíveis. Na arquitetura de análise moderna, bancos de dados NoSQL estão se tornando parte do arsenal de todas as organizações devido às vantagens da possibilidade de carregar rapidamente dados de qualquer fonte, incluindo fontes de dados sem esquemas ou formatos bem definidos. Os bancos de dados NoSQL, também chamados de "não SQL" ou "não apenas SQL", oferecem tipos alternativos de armazenamento de dados em comparação com os bancos de dados relacionais tradicionais, incluindo tipos de armazenamento em colunas, documentos, chave/valor e gráficos.

O conceito de Big Data e lagos de dados está relacionado aos dados desestruturados. Dados são gerados em toda parte, às vezes em lugares aleatórios, e coletar todos eles e convertê-los em um formato utilizável pode ser uma tarefa frustrante. Tecnologias foram desenvolvidas para permitir que as ferramentas de análise se conectem aos dados brutos em seu estado original, em vez de obrigá-los a enquadrar-se em um formato específico primeiro.

Uma delas é chamada "lago de dados", que é um repositório de armazenamento capaz de conter grandes volumes de dados em seu formato nativo, estruturados ou não. Assim, as pessoas podem analisar os dados usando mecanismos de processamento otimizados, como APIs e linguagens semelhantes a SQL, para transformar os dados conforme necessário, sem a necessidade de pré-processar todos eles em um formato específico.

Todas essas ferramentas geralmente são usadas em projetos relacionados à Internet das Coisas, à ciência dos dados, a dados em fluxo contínuo e a outros casos de uso de análise de dados desestruturados em que a criação de dados é imprevisível tanto em termos de volume como de localização.

Para ver uma lista das tecnologias de dados relacionadas ao padrão NoSQL, ao Hadoop e a lagos de dados, consulte o [Apêndice](#).

Arquivos sem formatação

A popularidade do Excel e dos arquivos CSV está longe de chegar ao fim. Não importa se sua organização é de grande ou pequeno porte, esses arquivos sem formatação que parecem ter geração espontânea continuarão existindo, provavelmente para sempre. Na realidade, eles estão espalhados por mais lugares do que nunca. Antes eles existiam apenas no computador físico de um usuário. Agora, habitam sistemas de armazenamento em nuvem, como o Google Drive e o Dropbox. As organizações criam arquivos sem formatação como parte dos "dados como serviço". Eles são úteis como legendas para vários campos de dados, pesquisas adicionais sobre clientes ou qualquer informação simples que amplie um conjunto de dados existente, pois podem ser criados rapidamente.

Além disso, aplique as medidas de segurança adequadas aos arquivos sem formatação no momento certo. Incentive seu uso quando necessário, principalmente em casos pontuais. Se alguns arquivos se tornarem populares, aplique os protocolos de segurança adequados, garantindo que eles estejam protegidos e só possam ser acessados pelas pessoas certas.

Criação e consumo

A marca distintiva do business intelligence moderno é a disponibilização da plataforma de business intelligence para os usuários corporativos. Acabaram os dias em que um tomador de decisões precisava solicitar um relatório à equipe de TI e esperar vários dias para receber um relatório desatualizado que não respondia à pergunta original com precisão. Hoje, o próprio tomador de decisões pode usar as ferramentas para responder sua pergunta por conta própria. Como a equipe de TI garantiu que toda a organização possa confiar nos dados disponíveis, os usuários corporativos podem tomar decisões inteligentes impulsionadas por dados conforme necessário, sem conhecimentos de programação.

O componente integrante essencial neste caso é a ferramenta de análise em si, algo que você pode começar a usar antes de qualquer coisa. Antes mesmo de identificar quais dos componentes integrantes apresentados nesta seção são os mais relevantes para sua organização, você deve começar a conectar-se aos seus dados onde quer que eles estejam para validar e explorar suas fontes de dados à medida que se prepara para desenvolver sua arquitetura completa.

Embora existam vários componentes para analisar os dados, acreditamos que o principal seja a análise visual, uma abordagem que permite que qualquer pessoa na sua organização (não importa se ela tem ou não conhecimentos de programação) se conecte diretamente a uma fonte de dados e extraia informações dela. Ao colocar esse tipo de ferramentas nas mãos de professores, médicos e membros da equipe de vendas, sua organização deixará de ser uma acumuladora de solicitações de alteração para se tornar uma máquina funcionando a todo vapor.

Uma novidade na área de business intelligence é o maior foco nos métodos de compartilhamento de informações com outros usuários. As pessoas não estão mais presas aos painéis e relatórios. Elas agora podem criar aplicativos interativos completos, longos artigos que se desenrolam em dados, textos e imagens ou até mesmo experiências visuais otimizadas para dispositivos móveis.

Além disso, à medida que as empresas e seus departamentos menores crescem, eles recorrem a ferramentas de produtividade inteligentes para compartilhar informações com rapidez, descobrir fontes de dados, acompanhar atualizações em painéis e seguir as métricas mais importantes.

Nesta seção, examinaremos alguns dos componentes que as melhores ferramentas de análise moderna oferecem para impulsionar o fornecimento dos dados. Para conferir mais conceitos, como histórias e alertas, consulte o [Apêndice](#).

Análise visual

O sistema visual humano é uma das ferramentas mais poderosas do mundo. Ele finalmente se tornou uma parte integrante da análise de dados. Aplicando o reconhecimento de padrões que o cérebro já utiliza diariamente, a análise visual pode revelar padrões nos dados de forma semelhante, como tendências de aumento ou diminuição, picos irregulares de atividade ou registros específicos que representam exceções.

Com as planilhas tradicionais, era necessário analisar os dados em linhas e colunas, selecionar um subconjunto a ser compartilhado e depois criar um gráfico. Tanto com assistentes ineficientes ou comandos de texto, esses gráficos às vezes respondiam à pergunta e às vezes levantavam outras perguntas totalmente novas, mas sempre eram o último passo que terminava em um beco sem saída. Em contrapartida, a análise visual proporciona uma experiência de análise elegante e intuitiva com simples ações de arrastar e soltar que tornam os elementos visuais parte do processo, oferecendo mais do que um simples gráfico como resultado final: a jornada de descoberta de informações é tão valiosa quanto a resposta.

A análise visual não se limita a lindas ferramentas de visualização. Trata-se de uma linguagem que permite combinar dados, identificar anomalias e expandir os dados com cálculos, agrupamentos, condições hipotéticas e muito mais, tudo isso sem exigir conhecimentos de programação.

BI e relatórios tradicionais

Ainda há espaço para o BI, os painéis e os relatórios tradicionais, embora seu processo de criação agora seja diferente. Para criar vários relatórios estáticos, como painéis executivos ou auditorias financeiras, antes era necessário ter conhecimentos técnicos de desenvolvimento, e as consultas analíticas eram definidas com muita antecedência, sendo frequentemente necessário alterar os modelos de dados subjacentes. Todo esse processo podia levar dias, semanas ou até mesmo meses para ser realizado.

Em uma plataforma de análise moderna, muitos desses painéis e relatórios nascem como perguntas pontuais que, devido à natureza dos questionamentos que respondem, são certificadas e validadas pela equipe de TI e pelos administradores de dados e, por fim, substituem os relatórios estáticos tradicionais. Esse processo modernizado aproveita a especialização dos usuários corporativos, pois, à medida que eles analisam os dados para encontrar respostas, as perguntas evoluem, mudam e suscitam perguntas totalmente novas. A flexibilidade da análise moderna está substituindo as ferramentas tradicionais, ainda que a necessidade desses relatórios tradicionais continue existindo.

Ferramentas de preparação de dados pessoais

Não devendo ser confundidas com ETL (extração, transformação e carregamento), as ferramentas de preparação de dados são aplicativos leves desenvolvidos para ajudar usuários que não são do departamento de TI a manipular os dados de formas avançadas e precisas. Elas são criadas a partir dos princípios de facilidade de uso, velocidade e agilidade das ferramentas de análise visual, que permitem que usuários corporativos comuns combinem conjuntos de dados, automatizem uniões, renomeiem campos e façam outras melhorias nos dados a fim de deixá-los prontos para análise.



Segundo o artigo [“What’s Your Data Strategy?”](#) (Qual é a sua estratégia de dados) da Harvard Business Review, “os analistas gastam 80% do seu tempo apenas na descoberta e na preparação dos dados”

Elas devem ser incluídas mais perto do fim da sua lista de prioridades, pois geralmente é difícil saber como você precisará modificar os dados até que comece a utilizá-los. Porém, ao longo da vida de uma organização, grande parte do tempo dedicado a responder a uma pergunta é usada para converter os dados no formato correto. As ferramentas de preparação de dados pessoais são uma forma eficiente de reduzir o tempo gasto nessa tarefa, sem necessariamente envolver ou negar a necessidade de uma camada semântica de relatórios predefinidos criada pela TI.

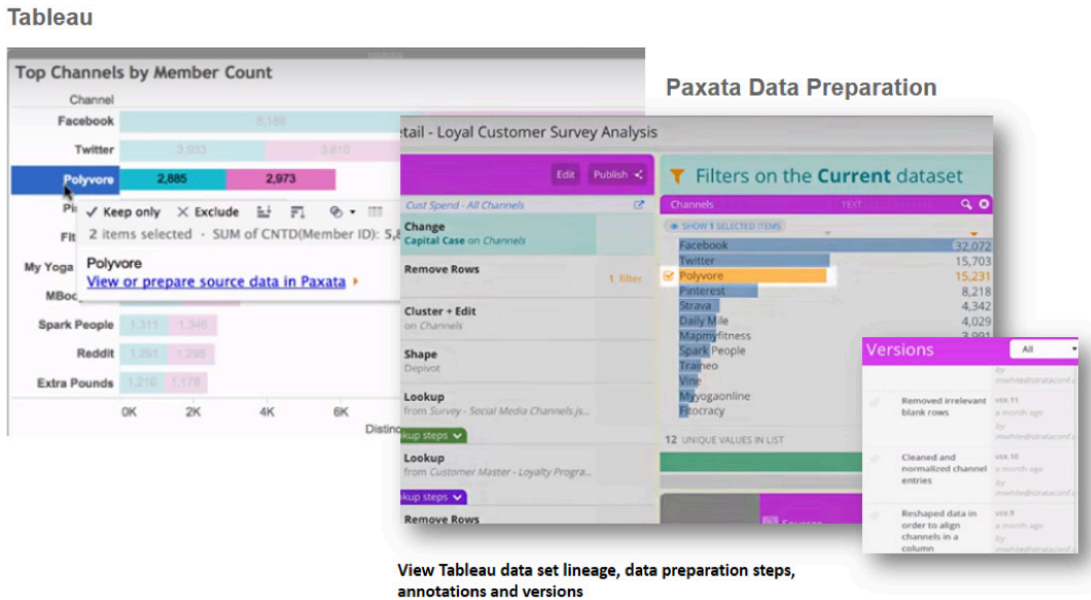


Figura 3 Integração da análise visual com a preparação de dados pessoais da Paxata

Análise avançada

A análise avançada surgiu como um componente essencial da arquitetura de análise moderna, já que as empresas recorrem a estatísticas, algoritmos preditivos e ao aprendizado de máquina para extrair o máximo valor de grandes conjuntos de dados.

Antigamente, a análise avançada só era acessível para cientistas de dados treinados que usavam programas como R, Python, SPSS ou SAS. Graças aos avanços na análise visual, hoje a análise avançada está ao alcance de todos os usuários da análise de autoatendimento na forma de recursos integrados. Recursos como caixas estreitas, mapas de árvore e técnicas básicas de modelagem preditiva são comuns e podem ser acessados com um clique do mouse.

Ainda há muitos casos de uso para as ferramentas dedicadas de análise estatística. Para organizações interessadas em criar algoritmos evolutivos para oferecer aos clientes "a próxima grande inovação" ou criar funções para identificar quando ocorrerem fraudes no cartão de crédito de um cliente, convém pesquisar ferramentas dedicadas específicas para complementar seu kit de ferramentas principais de análise visual. A implantação dessas ferramentas requer treinamento e conhecimentos de programação nas ferramentas específicas, algo que pode levar meses para ser desenvolvido.

Compartilhamento e colaboração

Nas plataformas de BI modernas, o compartilhamento, a colaboração e a publicação de informações são recursos fundamentais. Seja para avaliar o contexto ou priorizar a melhor forma de proceder, a colaboração maximiza os benefícios proporcionados pelas informações. As plataformas de análise modernas oferecem fóruns de discussão,

anotações, comentários, favoritos, curtidas e outros conceitos sociais adaptados dos principais aplicativos de portal e produtividade. A possibilidade de apresentar as informações descobertas em uma plataforma de análise diretamente na ferramenta simplifica muito o fluxo de análise e estimula os usuários a continuar explorando e discutindo sobre descobertas valiosas. Também é possível ampliar a colaboração para aplicativos e portais externos através da incorporação.

Análises incorporadas

Um dos conceitos empresariais mais eficientes e frequentemente ignorado é o fluxo. Em vez de tirar os usuários corporativos de seus processos operacionais comuns para procurar respostas nos dados, integre essas informações diretamente nos fluxos estabelecidos e incorpore-as nos processos existentes.

Com a análise moderna, você encontrará dados e painéis incorporados diretamente nos portais da empresa e em outros aplicativos ou integrados a ferramentas de produtividade. As melhores plataformas de análise contemplam esses casos de uso com APIs sofisticadas, SDKs e mecanismos de fornecimento flexíveis que permitem saltar de uma ferramenta à outra com facilidade e até mesmo unificam todas as suas ferramentas em um só portal.

O fluxo também se aplica à localização física. Os profissionais modernos frequentemente estão na estrada, sem acesso aos recursos internos cercados pelo firewall. A análise moderna permite acessar os dados em qualquer lugar, com qualquer tipo de dispositivo. Isso significa que um representante de vendas pode tomar uma decisão inteligente impulsionada por dados usando seu dispositivo móvel sem precisar ligar o laptop. Também significa que um gestor de obras que está na construção pode acessar informações importantes sem a necessidade de uma conexão de VPN com a rede corporativa. A tecnologia móvel e a nuvem mudaram para sempre a forma como as empresas podem executar suas operações, e plataformas de análise verdadeiramente modernas também devem possibilitar que as empresas aproveitem essas vantagens.

3. Testando todos os componentes em conjunto

Esses componentes integrantes são as bases de uma plataforma de análise moderna que permite à empresa enfrentar e superar qualquer desafio. Combinada a uma sólida parceria entre a empresa e a equipe de TI, essa plataforma proporciona a todos na sua empresa a confiança de que terão as ferramentas para tomar todas as decisões necessárias, e que essas decisões serão embasadas em dados confiáveis.

A simples tarefa de estabelecer uma plataforma de análise moderna completa pode parecer um desafio. A boa notícia é que você não precisa criar todo o ecossistema antes de começar. Na verdade, as empresas modernas são mais bem-sucedidas quando não fazem isso. Em vez disso, elas começam aos poucos e fazem transformações graduais que acabam revelando o próximo recurso no qual a empresa deve investir. As empresas podem começar com testes piloto antes de implementar as soluções de tecnologia em todos os departamentos. Você não precisa implementar todos os componentes da estratégia inteira para dar os primeiros passos. Por exemplo, você pode usar sua ferramenta de análise visual para identificar falhas no canal de dados antes de finalizar o warehouse. Assim, é possível extrair valor imediato da análise, encontrar possíveis lacunas e erros nos dados e, conseqüentemente, construir um data warehouse mais preciso e funcional.

O importante é trabalhar com ferramentas que possibilitem esse tipo de alterações graduais, e uma plataforma de análise moderna é justamente um conjunto de componentes integrantes que podem ser incorporados um de cada vez, tornando sua organização mais acessível, ágil e capaz de extrair informações de uma ampla variedade de fontes

de dados. É exatamente isso que os líderes mundiais em análise estão fazendo: empregando a análise visual de forma estratégica junto com outras soluções excelentes de análise de Big Data, Internet das Coisas e ciência de dados.

Por exemplo, [a Netflix criou uma plataforma de Big Data abrangente](#) e um lago de dados para comportar as imensas quantidades de dados geradas pelas operações da empresa. O Tableau é um componente essencial que permite à empresa combinar suas diversas ferramentas, como S3, EMR e Spark, em uma única plataforma coesa que contribui para os negócios.

Não importa em qual estágio do processo de transformação digital dos negócios você está: é fundamental começar imediatamente a usar os dados que você já tem. As empresas precisarão agir cada vez mais rápido diante da próxima grande revolução do mercado. Escolha um componente da análise moderna e capacite a empresa a tomar decisões impulsionadas por dados e a liderar a revolução.

4. Apêndice

Há muitas opções de soluções e tecnologias de análise diferentes, e cada uma delas tem finalidades e vantagens específicas. Neste Apêndice, detalharemos todas as opções que não foram abordadas na visão geral.

Processamento em fluxo contínuo

Os dados em fluxo contínuo são gerados continuamente por dispositivos conectados e aplicativos localizados em toda parte, como redes sociais, medidores inteligentes, automação residencial, videogames e sensores de IoT. Esses dados geralmente são coletados através de canais de dados semiestruturados. Embora a análise em tempo real e os algoritmos preditivos possam ser aplicados aos fluxos contínuos, os dados em fluxo contínuo geralmente são encaminhados e armazenados em formatos brutos através da [arquitetura Lambda](#) e em um lago de dados, como o Hadoop, para uso nas análises.

A Lambda é uma arquitetura de processamento de dados desenvolvida para manipular quantidades maciças de dados empregando métodos de processamento em lote e em fluxo contínuo. O conceito equilibra os desafios de latência, taxa de transferência e tolerância a falhas.

Diversas opções estão disponíveis atualmente para dados em fluxo contínuo, como Amazon Kinesis, Storm, Flume, Kafka e Informatica Vibe Data Stream.

Orquestração de hubs de integração

Os padrões de integração "hub-and-spoke" são um modelo de arquitetura de dados amplamente utilizado e fácil de entender. Os hubs desvinculam fontes de dados localizadas em qualquer lugar e visam possibilitar uma integração mais flexível reduzindo o número de interfaces "ponto a ponto" a serem gerenciadas. Os recursos de publicação/assinatura dos hubs de integração promovem a reutilização de dados e oferecem controle centralizado para fins de otimização, implementação de padrões de dados e governança. O gerenciamento centralizado aumenta a visibilidade de todos os canais de movimentação de dados orquestrados que englobam fontes de dados localizadas em qualquer lugar.

Os hubs de integração de dados de última geração disponibilizam os recursos tradicionais para os usuários da análise de autoatendimento. Qualquer pessoa pode publicar ou assinar feeds de dados de hubs de integração modernos com o mínimo de intervenção da equipe de TI. Os consumidores de dados podem aproveitar dados certificados, adquirir visibilidade da linhagem dos dados e dos processos de integração. Outras vantagens dos hubs de integração de dados modernos são as funções de qualidade de dados integradas, integração acelerada de fontes de dados e fornecimento sob demanda de conjuntos de dados pequenos ou imensos.

A Informatica e a Cisco são líderes de mercado em tecnologia de hub de integração de dados. Graças à profunda integração do Tableau com a Informatica, você pode combinar centenas de fontes de dados diferentes em extrações de dados do Tableau armazenadas e atualizadas no [Servidor de Dados do Tableau](#) para serem usadas por qualquer pessoa na organização.

Mais informações sobre dados desestruturados, NoSQL e lagos de dados

Os lagos de dados satisfazem as necessidades modernas de análise de Big Data com armazenamento e processamento de dados mais rápidos e flexíveis, permitindo que qualquer pessoa analise dados brutos de diversas formas com rapidez. Os lagos de dados não substituem os data warehouses.

Nos padrões de design modernos de "processamento e carregamento", o destino de dados brutos de qualquer tamanho ou formato geralmente é um lago de dados. Um lago de dados é um repositório de armazenamento que contém uma enorme quantidade de dados em seu formato nativo: estruturados, semiestruturados e desestruturados. Os lagos de dados também oferecem mecanismos de processamento otimizados com APIs e linguagens semelhantes a SQL para transformar os dados brutos com a funcionalidade de "esquema na leitura".

Embora o Hadoop venha sendo utilizado para lagos de dados desde o primeiro HDFS (Hadoop Distributed File System) por sua adaptabilidade e seu baixo custo, ele não é a única opção para implementar um lago de dados. Repositórios de objetos como o Amazon Web Services Simple Storage Service (S3) e bancos de dados NoSQL com esquemas flexíveis também podem ser utilizados como lagos de dados. O Tableau [agora é compatível com o serviço de dados Amazon Athena](#) para conectar-se ao Amazon S3 e tem várias ferramentas que permitem conectividade direta com bancos de dados NoSQL.

Nas arquiteturas de análise modernas, os bancos de dados NoSQL estão se tornando o padrão devido aos benefícios do rápido carregamento de dados em qualquer lugar e dos conceitos de bancos de dados sem esquema. Bancos de dados NoSQL, "não SQL" ou "não apenas SQL" oferecem tipos alternativos de armazenamento de dados. Tipos comuns de armazenamento NoSQL incluem colunas, documentos, valor/chave e gráficos.

Exemplos de bancos de dados NoSQL que geralmente são utilizados com o Tableau são MongoDB, Datastax e MarkLogic, entre outros.

Embora o Hadoop normalmente seja usado como uma plataforma de Big Data, ele não é um banco de dados. O Hadoop é uma estrutura de software de código aberto para armazenar dados e executar aplicativos em clusters de hardware comercial. Ele oferece armazenamento maciço para qualquer tipo de dados, imensa capacidade de processamento e a possibilidade de processar volumes gigantescos de tarefas e trabalhos simultâneos.

Em uma arquitetura de análise moderna, o Hadoop oferece armazenamento e arquivamento de dados de baixo custo para transferir dados históricos antigos do data warehouse para repositórios inativos on-line. Ele também é utilizado para casos de uso de análise de dados desestruturados, IoT e ciência de dados.

Na estrutura do Hadoop, as tecnologias relacionadas de carregamento, organização e consulta de dados incluem, entre outras, as seguintes:

- **Apache Spark:** estrutura de computação em cluster de código aberto com análise na memória de alto desempenho e um número crescente de projetos relacionados
- **Apache Impala:** banco de dados MPP analítico de código aberto para Apache Hadoop. Essa é a conexão de dados mais usada em projetos bem-sucedidos do Hadoop no Tableau
- **Apache Presto:** um mecanismo de consultas SQL distribuído de código aberto para executar consultas interativas em vários conjuntos de dados de todos os tamanhos. [A versão 10 do Tableau incluiu suporte ao Presto](#)
- **MapReduce:** uma estrutura de software de processamento paralelo que recebe entradas, as segmenta em problemas menores e as distribui para os nós de trabalho

- **Hive:** um sistema de data warehouse e uma linguagem de consulta semelhante a SQL. O Hive 2.0 também inclui o LLAP (Live Long and Process), que melhora significativamente o desempenho de consultas do Hive.
- **HDFS (Hadoop Distributed File System):** sistema escalonável que armazena dados em várias máquinas sem organização prévia
- **YARN (Yet Another Resource Negotiator):** oferece gerenciamento de recursos para processos em execução no Hadoop
- **Ambari:** uma interface na Web para gerenciar os serviços e componentes do Hadoop
- **Cassandra:** um sistema de banco de dados distribuído
- **Flume:** software para transmitir dados para o HDFS
- **HBase:** um banco de dados distribuído não relacional que é executado no Hadoop
- **HCatalog:** uma camada de gerenciamento de tabelas e armazenamento
- **Oozie:** um agendador de tarefas do Hadoop
- **Pig:** uma plataforma para manipular dados armazenados no HDFS
- **Solr:** uma ferramenta de pesquisa escalonável
- **Sqoop:** transfere dados entre o Hadoop e bancos de dados relacionais
- **Zookeeper:** um aplicativo que coordena o processamento distribuído

Nos últimos dois anos em particular, o Apache Spark deixou de ser um componente do ecossistema do Hadoop para se tornar a plataforma de análise de Big Data independente adotada por muitas empresas. O Spark oferece uma velocidade de processamento de dados consideravelmente mais alta em relação ao Hadoop. O Spark tem muitos projetos relacionados, incluindo o tempo de execução principal do Apache Spark, Spark SQL, Spark Streaming, MLlib, ML e GraphX. Ele é hoje o maior projeto de Big Data de código aberto, com mais de 1.000 contribuidores de mais de 250 organizações.

A Tableau é uma líder de mercado em análise de dados visual e conectividade específica para a análise de Big Data. Os melhores programas de análise de Big Data usam o Tableau com as plataformas Cloudera, Spark SQL, Amazon EMR, Hortonworks, Microsoft HDInsight/Data Lake e MapR. Muitas outras tecnologias de Big Data podem ser conectadas ao Tableau através dessas tecnologias abertamente compatíveis ou de seus drivers.

Dados como serviço

Em um mundo digital onde os dados são ouro, eles também são um produto que qualquer pessoa pode consumir. Dados financeiros, geográficos, demográficos, sobre clientes, mercados e previsão do tempo já estão à venda como serviços em mercados de dados e plataformas de comércio.

Os dados como serviço aplicam um padrão SOA (arquitetura orientada a serviços) flexível para fornecimento de dados através da nuvem. Essa abordagem oferece máxima agilidade, pois a arquitetura SOA é extremamente simples. Hoje vemos ISVs, CRMs e ERPs que oferecem APIs REST padrão de dados como serviço para fins de integração ou de criação de relatórios externos.

Com o [SDK do Conector de dados da Web do Tableau](#), as pessoas podem estabelecer conexões com dados situados fora dos conectores existentes. Os usuários da análise de autoatendimento podem se conectar a praticamente todos os dados acessíveis via HTTP, incluindo serviços Web internos, dados JSON e APIs REST.

Data warehouse lógico

Organizações líderes em análise estão oferecendo exibições de dados flexíveis, lógicas, unificadas e dimensionais em qualquer lugar através de tecnologias de virtualização de dados de fornecedores como a Cisco e a Denodo. Para os usuários da análise, um data warehouse lógico se parece e funciona como um data warehouse relacional. Os usuários do Tableau podem conectar-se a essas ferramentas usando drivers ODBC predefinidos.

Um dos principais recursos da virtualização de dados é a otimização de diversas consultas remotas distribuídas

em uma variedade de fontes de dados e APIs REST diferentes. O data warehouse lógico também serve como uma camada semântica que protege os aplicativos de relatório contra alterações nas fontes de dados. Os data warehouses lógicos geralmente são utilizados com um catálogo de dados empresarial.

Gerenciamento de dados mestres

A eficácia da análise depende da qualidade dos dados utilizados, e as pessoas só podem tomar as decisões certas se os dados forem precisos. Como mais especialistas estão criando dados com uma abordagem "de baixo para cima", estamos observando o ressurgimento do interesse pela qualidade dos dados tradicionais e pelo gerenciamento de dados mestres, que estão voltando a ser fundamentais para garantir que as fontes de dados de relatórios sejam atuais, corretas, consistentes e precisas.

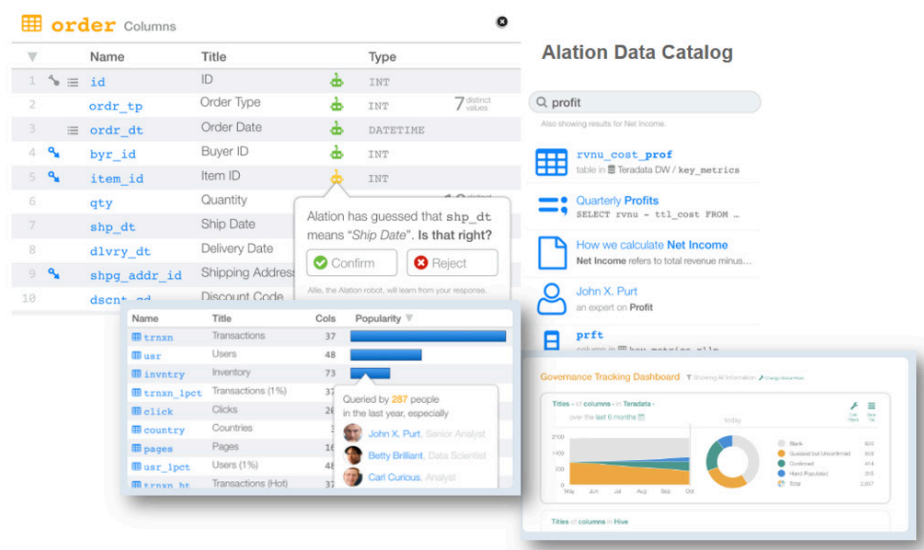
Produtos populares de gerenciamento de dados mestres são oferecidos por empresas como a Informatica, a IBM e a Stibo, entre outras. Várias soluções de qualidade de dados frequentemente utilizadas com o Tableau são: Trillium, Informatica Data Quality, Talend Data Quality 6.0 e Tamr Eisenhower.

Catálogo de dados empresarial

Outra tecnologia emergente é o catálogo de dados empresarial. Catálogos de dados empresariais permitem que usuários de relatório independentes rapidamente encontrem os dados certos para tomar decisões utilizando fontes de dados aprovadas. Os catálogos de dados empresariais fazem parte das soluções de análise visual e também estão disponíveis como produtos independentes que se integram diretamente ao Tableau.

Os catálogos de dados empresariais são preenchidos com metadados de tabelas, exibições e procedimentos armazenados através de varreduras nas fontes de dados processadas. Com a descoberta automatizada de novas fontes de dados, a classificação inteligente de dados e o mapeamento de entidades entre fontes de dados, um catálogo de dados basicamente serve como um glossário das fontes de dados e definições de dados comuns de uma empresa. Os especialistas técnicos melhoram ainda mais o contexto das fontes de dados do catálogo adicionando anotações, versões e documentação.

Figura 4 Um catálogo de dados empresarial da Alation



As soluções de catálogo de dados promovem a manutenção dos dados e a reutilização eficiente dos dados existentes. Elas também oferecem uma linhagem de dados extremamente útil e um nível superior de governança de dados,

segurança, registros e auditoria.

Alguns fornecedores com sólidos catálogos de dados que se integram perfeitamente ao Tableau incluem: Alation, Collibra, Attivio, Informatica e Waterline.

Aprendizado de máquina

Ampliando ainda mais o potencial da análise avançada, o aprendizado cognitivo aprofundado e a inteligência artificial fazem deduções com base nos dados e padrões existentes, tiram conclusões fundamentando-se nas bases de conhecimento disponíveis e reinserem essas informações em uma base de conhecimento em ciclos de autoaprendizado contínuos infinitos.

O uso desse tipo de análise geralmente pressupõe examinar os resultados através de uma API de incorporação em um relatório ou aplicativo integrado. O Tableau é utilizado hoje para visualizar os resultados da CognitiveCode, da Digital Reasoning e de outros fornecedores.

No Tableau 10.3, lançamos as [tabelas recomendadas e uniões inteligentes](#) para que você possa economizar tempo ao conectar-se aos seus dados e prepará-los. Com o uso do aprendizado de máquina, as recomendações melhoram ao longo do tempo à medida que os bancos de dados são utilizados com cada vez mais frequência.

Linguagem natural

Ampliando ainda mais os benefícios de contar histórias com os dados, a linguagem natural e a consulta por fala estão começando a democratizar a descoberta de dados de formas novas e mais flexíveis. A linguagem natural torna a análise mais acessível através de qualquer plataforma obtendo descrições contextuais automatizadas de descobertas relevantes, solicitando previsões ou analisando grandes volumes de documentos de texto.

Hoje, a análise visual do Tableau pode ser combinada com soluções líderes de NLG (geração de linguagem natural), como Yseop, Narrative Science e Automated Insight. Como essas tecnologias interpretam amplamente o contexto das visualizações do Tableau, as integrações quase sempre são realizadas na própria ferramenta de linguagem natural, ou como uma extensão por meio de JavaScript. Além disso, a [aquisição da ClearGraph](#) permitirá análises e descobertas de dados mais inteligentes no Tableau, facilitando a interação com os dados através da linguagem natural.

Descoberta de dados recomendada

A descoberta de dados inteligente emprega algoritmos de aprendizado de máquina para oferecer recursos analíticos ("o que acontecerá") e prescritivos ("como otimizar") muito mais aprofundados com os dados formatados para essa finalidade. À medida que a análise visual evolui, novas informações e recomendações automatizadas serão incorporadas. Esses recursos foram mostrados na palestra de roteiro apresentada na [Conferência da Tableau 2016](#). Integrando R, Python, APIs e funções de banco de dados analíticas, os resultados dos recursos de análise avançada podem ser visualizados e explorados no Tableau.

Pesquisa

As arquiteturas de análise modernas permitem que os usuários e especialistas pesquisem e encontrem dados com a mesma facilidade de usar o Google, independentemente de onde eles estejam localizados. Em vez de modelar os dados, a tecnologia de indexação de mecanismo de pesquisa analítica relaciona diferentes fontes de dados de acordo com os nomes de campos, tipos de dados e conclusões do aprendizado de máquina. Com o passar do tempo, sugestões de pesquisa dinâmicas são geradas com base em consultas anteriores e no uso registrado. Recentemente, com a incorporação de tecnologias de fala, como Siri e Alexa, estamos começando a ver recursos

de consulta por voz combinados com a pesquisa analítica. Um ótimo exemplo foi o [vencedor da Hackathon anual](#) na Conferência da Tableau 2016 da Automated Insight, que integrou o controle de voz da Alexa ao Tableau.

Alertas

Uma arquitetura de análise moderna traz alertas de notificação configuráveis, inteligentes e acionados por dados que monitoram continuamente o oceano digital de dados em busca de sinais relevantes. É humanamente impossível verificar todos os valores importantes 24 horas por dia, 7 dias por semana. É por isso que a automação e os alertas são recursos maravilhosos do arsenal da análise moderna.

Algumas ferramentas apresentam resumos em intervalos regulares. Outras monitoram os registros para verificar se os números ultrapassaram determinados limites. Ambos os recursos têm sua utilidade. Alguns painéis são informativos e você deseja consultá-los diariamente. Outros servem como referência para ações importantes, mas consultar diariamente um painel sem informações úteis é um desperdício de tempo.

No Tableau, você pode manter-se atualizado sobre seus negócios com [alertas acionados por dados para o Tableau Server](#). Basta definir os critérios para que um alerta seja enviado por e-mail para você ou para toda a equipe.

Histórias

Às vezes, a informação ("o quê") não é suficiente. As pessoas também querem saber o "porquê" por trás dos dados. Por que as vendas aumentaram? O que provocou o pico de tráfego no site? Por que estamos com dificuldade para manter o estoque de suprimentos médicos?

Há muito tempo, as empresas tentam resolver esse problema combinando a análise com outras formas de comunicação: texto, imagens e até mesmo vídeos. Os analistas criam apresentações usando o PowerPoint, redigem longos relatórios em formato PDF ou até mesmo imprimem e encadernam documentos com um incontável número de páginas.

As ferramentas de análise moderna identificam os melhores conceitos da criação de histórias e os incorporam como recursos principais. Elas permitem criar painéis interativos, enviar instantâneos específicos dos dados que são automaticamente atualizados em segundo plano quando novos dados são adicionados ou até mesmo elaborar relatórios que combinam gráficos interativos com texto e imagens. Os usuários podem contar histórias para explicar a análise dos dados em vez de simplesmente apresentar números.

Por princípio, o Tableau valoriza a importância da liberdade de escolha e dos padrões abertos. Investimos muito em pesquisas e desenvolvimento para acelerar e simplificar a análise, e isso também significa inovar com nosso ecossistema de parceiros. Isso garante que, à medida que o mundo da análise evoluir e novas tecnologias surgirem, os líderes em análise sempre possam integrar o Tableau às tecnologias de dados atuais e futuras de sua preferência.

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. [Avalie o Tableau gratuitamente](#) hoje mesmo!

Recursos adicionais

[Como desenvolver uma cultura de análise](#)

[Definição de análise](#)

[Abordagem de análise \(página Soluções\)](#)

[Resumo sobre os dados na nuvem](#)

[Análises avançadas com o Tableau](#)

