

10 PRINCIPAIS
TENDÊNCIAS DE
Big Data





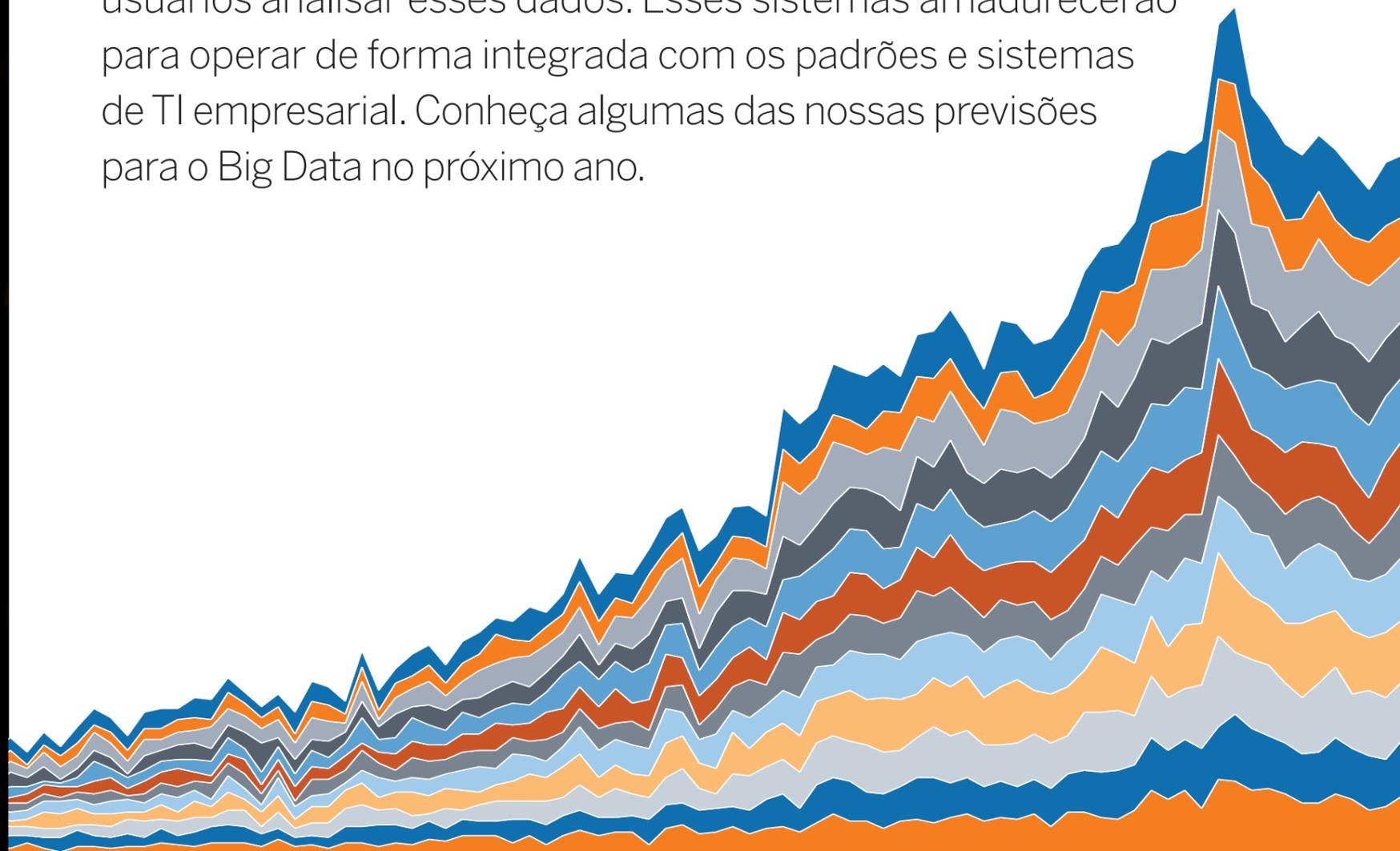
Dez principais tendências de Big Data

Mais organizações estão armazenando, processando e extraindo valor de dados de todos os tipos e tamanhos. Os sistemas que oferecem suporte a grandes volumes de dados estruturados e não estruturados continuarão crescendo. Haverá uma demanda de mercado por plataformas que ajudem os administradores de dados a governar e proteger o Big Data e que permitam aos usuários analisar esses dados. Esses sistemas amadurecerão para operar de forma integrada com os padrões e sistemas de TI empresarial. Conheça algumas das nossas previsões para o Big Data no próximo ano.

Todos os anos na Tableau, conversamos sobre o que está acontecendo no setor.

A partir dessa discussão, criamos a lista das principais tendências para Big Data do ano seguinte.

Estas são as nossas previsões.



BIG DATA

1

Big Data mais ágil e acessível: mais opções para agilizar o Hadoop

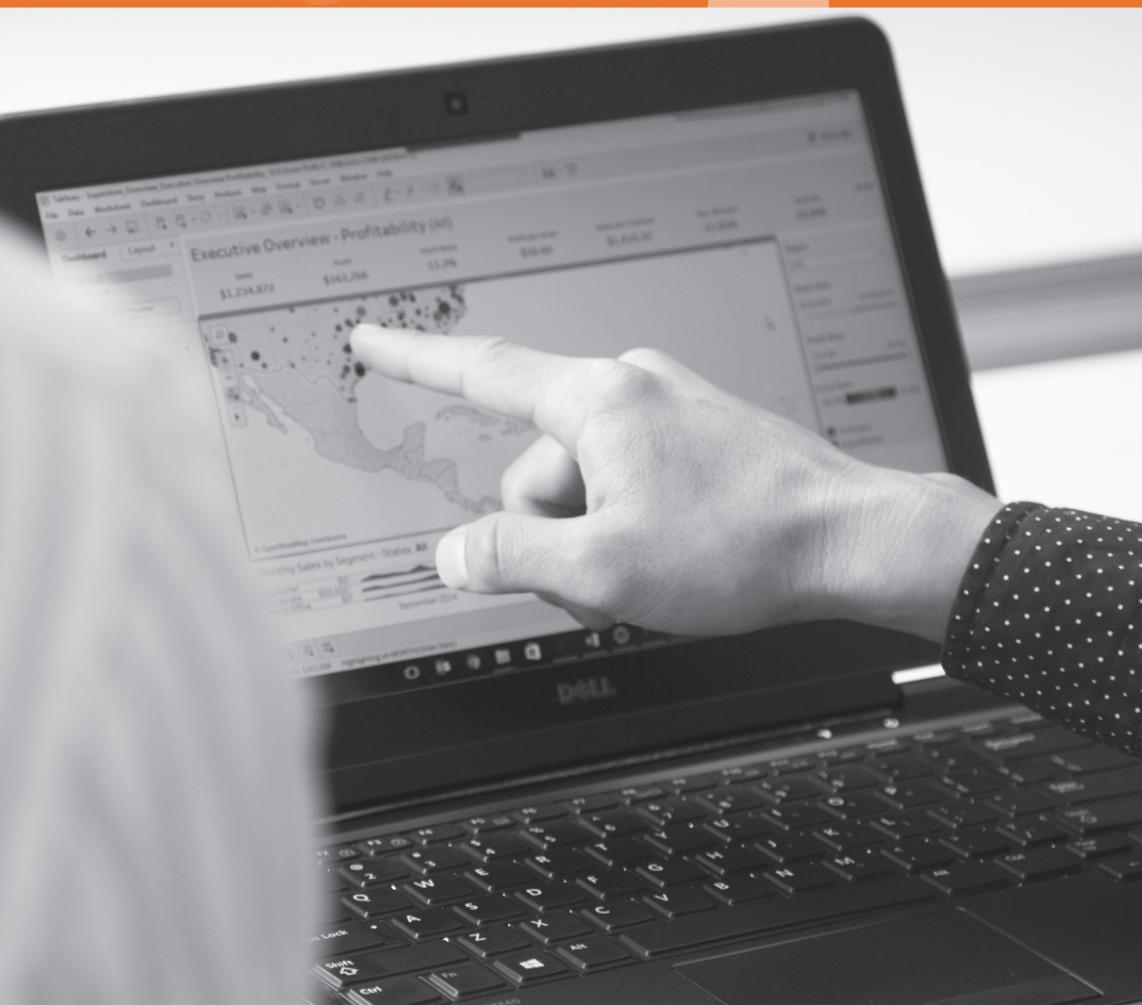
É claro que você pode utilizar o aprendizado de máquina e fazer análises de sentimentos no Hadoop, mas a primeira pergunta que as pessoas geralmente fazem é: o quanto o SQL interativo é rápido? Afinal, o SQL é a via utilizada pelos usuários corporativos que desejam acessar os dados do Hadoop para criar painéis com KPIs mais iterativos e ágeis, bem como para fazer análises exploratórias.

Essa necessidade de rapidez incentivou a adoção de bancos de dados mais ágeis, como o [Exasol](#) e o [MemSQL](#), de armazenamentos baseados no Hadoop, como o [Kudu](#), e de tecnologias que possibilitam consultas mais rápidas. Com a utilização de mecanismos SQL no Hadoop ([Apache Impala](#), [Hive LLAP](#), [Presto](#), [Phoenix](#) e [Drill](#)) e de tecnologias que utilizam o OLAP no Hadoop ([AtScale](#), [Jethro Data](#) e [Kyvos Insights](#)), esses aceleradores de consulta estão tornando ainda mais tênue a linha que separa os warehouses tradicionais do mundo do Big Data.

O Big Data não está mais limitado ao Hadoop: ferramentas específicas para o Hadoop estão ficando obsoletas

Nos anos anteriores, com a onda do Big Data, várias tecnologias surgiram para atender às necessidades de análises no Hadoop. No entanto, empresas com ambientes complexos e heterogêneos não querem mais adotar um ponto de acesso de BI específico apenas para um tipo de fonte de dados (Hadoop). As respostas para as perguntas dessas empresas estão escondidas em diversas fontes de dados, que variam desde sistemas de registros até dados estruturados e não estruturados de fontes de dados do Hadoop e de outros tipos. (Por sinal, até mesmo os bancos de dados relacionais já estão se preparando para processar Big Data. O SQL Server, por exemplo, adicionou recentemente o suporte a JSON.)

Os clientes precisarão fazer análises com todos os tipos de dados. Consequentemente, as plataformas que são agnósticas em relação a dados e fontes prosperarão, enquanto aquelas desenvolvidas especificamente para o Hadoop e que não são compatíveis com outros casos de uso serão deixadas de lado. A [aquisição da Platfora](#) é um indicador dessa tendência.





Organizações aproveitam lagos de dados desde o início para obter valor

Um lago de dados é como um reservatório. Primeiro você cria a estrutura (um cluster) e depois enche de água (dados). Depois que o lago estiver pronto, você começa a usar a água (dados) para várias finalidades, como geração de energia, consumo e recreação (análises preditivas, aprendizado de máquina [ML], segurança cibernética, etc.).

Até aqui, alimentar o lago tem sido suficiente. Isso mudará, uma vez que não será mais tão fácil justificar comercialmente o uso do Hadoop. As organizações precisarão usar o lago repetidamente e com agilidade para obter respostas com mais rapidez. Elas considerarão cuidadosamente os resultados comerciais antes de investir em contratações, dados e infraestrutura. Isso favorecerá uma parceria sólida entre a **empresa e a TI**. Além disso, as plataformas de autoatendimento serão reconhecidas como as ferramentas ideais para lidar com ativos de Big Data.

4

Arquiteturas desenvolvidas para rejeitar um tamanho padrão para todas as estruturas

O Hadoop não é mais apenas uma plataforma de processamento em lotes para casos de uso de ciência dos dados. Ele se tornou um mecanismo multifuncional para análises ad hoc e tem sido usado até mesmo para gerar relatórios operacionais sobre as cargas de trabalho diárias; o que geralmente é feito por data warehouses.

As organizações responderão a essas necessidades híbridas buscando uma arquitetura desenvolvida especificamente para um caso de uso. Elas pesquisarão diversos fatores, incluindo o comportamento dos usuários, perguntas, volumes, frequência de acesso, velocidade dos dados e nível de agregação, antes de escolher uma estratégia de dados. Essas arquiteturas modernas serão desenvolvidas para atender a necessidades específicas. Elas combinarão as melhores ferramentas de preparação de dados de autoatendimento, o Hadoop Core e plataformas de análise voltadas para o usuário final, de forma a possibilitar sua reconfiguração à medida que essas necessidades evoluem. A flexibilidade dessas arquiteturas orientará as escolhas de tecnologia.



5

A variedade, e não o volume ou a velocidade, é o que orienta os investimentos em Big Data

A [Gartner](#) define o Big Data como os três Vs: volume alto, velocidade alta e variedade ampla de ativos de informação. Embora os três Vs estejam crescendo, a variedade orienta cada vez mais os investimentos em Big Data, como foi visto em uma [pesquisa recente](#) feita pela New Vantage Partners. Essa tendência continuará crescendo com as empresas, que buscam integrar mais fontes e se concentrar na “[cauda longa](#)” do Big Data. Do JSON sem esquema até tipos aninhados em outros bancos de dados (relacionais e NoSQL) e dados com formatação (Avro, Parquet, XML), os formatos de dados estão se multiplicando, e os conectores estão se tornando essenciais. Empresas continuarão avaliando as plataformas de análise com base em sua capacidade de oferecer uma conectividade direta em tempo real com essas diversas fontes.

O Spark e o aprendizado de máquina dão gás ao Big Data

O [Apache Spark](#), anteriormente um componente do ecossistema do Hadoop, está se tornando a plataforma preferida de Big Data das empresas. Em uma [pesquisa](#) com arquitetos de dados, gerentes de TI e analistas de BI, aproximadamente 70% dos entrevistados preferiam o Spark ao tradicional MapReduce, que é baseado em lote e não pode ser usado com aplicativos interativos ou no processamento de fluxo em tempo real.

Esses recursos de processamento de Big Data provocaram uma evolução nas plataformas, que agora oferecem aprendizado de máquina intensivo, IA e algoritmos de gráfico. O aprendizado de máquina do Microsoft Azure, em particular, emplacou graças à sua interface simples de usar e facilidade de integração com plataformas Microsoft existentes. Disponibilizar o aprendizado de máquina para as massas resultará na criação de mais modelos e aplicativos que, por sua vez, gerarão petabytes de dados. À medida que as máquinas aprendem e os sistemas ficam mais inteligentes, todos os olhares estarão voltados para os provedores de software de autoatendimento para ver como eles tornarão esses dados acessíveis para os usuários finais.

A convergência da Internet das coisas (IoT), da nuvem e do Big Data gera novas oportunidades para a análise de autoatendimento

Tudo terá um sensor que envia informações para a “nave-mãe”. A IoT está gerando volumes enormes de dados estruturados e não estruturados, e uma parte cada vez maior desses **dados está sendo implantada em serviços de nuvem**. Normalmente, os dados são heterogêneos e estão armazenados em diversos sistemas relacionais e não relacionais, desde clusters do Hadoop até bancos de dados NoSQL. Embora as inovações em armazenamento e serviços gerenciados tenham agilizado o processo de captura, acessar e entender os dados ainda é um desafio para o usuário final. Por isso, há uma crescente demanda por ferramentas de análise capazes de se conectar a uma ampla gama de fontes de dados hospedadas na nuvem e combiná-las. Ferramentas como essas permitem que as empresas explorem e visualizem quaisquer tipos de dados armazenados em qualquer lugar, ajudando-as a descobrir oportunidades escondidas em seus investimentos em IoT.

A preparação de dados de autoatendimento está se tornando essencial, à medida que os usuários começam a modelar o Big Data

Disponibilizar os dados do Hadoop para usuários corporativos é um dos maiores desafios do nosso tempo. O surgimento das plataformas de análise de autoatendimento facilitou essa jornada, mas os usuários corporativos querem reduzir ainda mais a complexidade e o tempo destinado à preparação de dados para análise, o que é especialmente importante quando há vários tipos e formatos de dados.

As ferramentas de preparação de dados de autoatendimento permitem que os dados do Hadoop sejam preparados na origem, além de disponibilizar os dados como instantâneos para facilitar e agilizar sua exploração. Vimos muitas inovações nesse campo, como o [Alteryx](#), o [Trifacta](#) e o [Paxata](#). Essas ferramentas estão reduzindo os obstáculos que impedem a entrada dos [clientes tardios](#) e [retardatários do Hadoop](#), e elas continuarão a ganhar espaço.

O Big Data cresceu: Hadoop reforça os padrões empresariais

Estamos observando uma tendência crescente do Hadoop como um elemento essencial do cenário de TI empresarial. Os investimentos em componentes de segurança e governança para sistemas empresariais aumentarão. O Apache Sentry oferece um sistema de autorização granular baseada em função para os dados e metadados armazenados em um cluster do Hadoop. O [Apache Atlas](#), desenvolvido como parte da iniciativa de governança de dados, permite que as organizações apliquem uma classificação de dados consistente a todo o ecossistema de dados. O [Apache Ranger](#) fornece um ambiente de gerenciamento de segurança centralizado para o Hadoop.

Esses são os tipos de recursos que os clientes esperam de suas plataformas RDBMS empresariais. Esses recursos agora estão no centro das mais novas tecnologias de Big Data, eliminando, assim, mais um obstáculo para que as empresas os adotem.

O surgimento de catálogos de metadados ajuda as pessoas a encontrar Big Data que vale a pena ser analisado

Por muito tempo, empresas descartaram seus dados porque não tinham capacidade de processar tantos deles. Com o Hadoop, elas conseguem processar grandes volumes de dados, mas nem sempre esses dados estavam organizados de uma forma propícia para sua descoberta.

Os catálogos de metadados podem ajudar usuários a descobrir e entender dados relevantes para a análise com o auxílio de ferramentas de autoatendimento. Essa lacuna na necessidade do cliente está sendo preenchida por empresas como a [Alation](#) e a [Waterline](#), que usam aprendizado de máquina para automatizar o trabalho de descoberta de dados no Hadoop. Elas catalogam arquivos usando marcas, descobrem relações entre ativos de dados e até mesmo fornecem sugestões de consultas em interfaces do usuário pesquisáveis. Isso ajuda os consumidores e os administradores de dados a reduzir o tempo destinado à confiabilidade, à descoberta e à consulta adequada dos dados. No próximo ano, haverá mais conscientização e demanda para a descoberta de autoatendimento, que crescerá como uma extensão natural da análise de autoatendimento.



Sobre a Tableau

Integrar a visualização de dados em seus programas e processos de varejo é mais fácil do que você pensa.

A Tableau Software ajuda as pessoas a ver e entender os dados independentemente do seu volume ou do número de sistemas nos quais eles estão armazenados. Conecte, combine, visualize e compartilhe painéis de dados num piscar de olhos com uma experiência integrada do PC ao iPad. Crie e publique painéis com atualizações de dados automáticas e compartilhe-os com colegas, parceiros ou clientes, sem precisar ter conhecimentos de programação. Inicie uma avaliação gratuita hoje mesmo.

[TABLEAU.COM/PT-BR/TRIAL](https://tableau.com/pt-br/trial)