

Oito maneiras de modernizar seu gerenciamento de dados

Como aproveitar a inteligência artificial, lidar com a crescente diversidade de dados e melhor integrar as implantações





Índice

Introdução

Os dados não são mais como eram antes

Página 3

Um

Impulsione a otimização e o desenvolvimento das aplicações com a IA

Página 4

Dois

Permita relatórios e análises rápidos

Página 6

Três

Reduza seu custo total de propriedade

Página 8

Quatro

Obtenha flexibilidade para escolher na nuvem ou localmente, de acordo com a sua necessidade

Página 10

Cinco

Aumente a escalabilidade e a disponibilidade constantes

Página 12

Seis

Ofereça suporte a novos tipos e fontes de dados

Página 14

Sete

Integre a tecnologia de código aberto

Página 16

Oito

Incorpore novas capacidades de streaming e fast data

Página 19

Conclusão

Como começar a modernizar seu gerenciamento de dados

Página 21

Introdução

Os dados não são mais como eram antes



Seu gerenciamento de dados também não deve ser

A época em que os dados tinham uma definição estreita, como informações altamente estruturadas de algumas fontes específicas, já terminou há muito tempo. Substituindo essa noção, está a realidade das organizações que procuram aproveitar uma variedade maior de tipos e estruturas de dados provenientes de múltiplas fontes. Tudo isso serve para fornecer a todos, desde TI a funcionários de linha de negócios (LOB) e executivos de nível C, insights que podem ter um impacto imediato e transformador.

Usar e gerenciar as informações efetivamente é fundamental para buscar novas oportunidades de negócios, atrair e reter clientes e otimizar as operações. No entanto, essas necessidades criam uma série de desafios de carga de trabalho e aumentam as demandas sobre a infraestrutura de TI e nos sistemas de banco de dados subjacentes que frequentemente não estão à altura da tarefa.

A questão é: como você resolverá esses desafios? Você alocará mais funcionários para acompanhar os patches, complementos e ajustes contínuos exigidos pelos sistemas existentes, ou simplesmente ignorará os possíveis insights que estão nesse enorme conjunto de novos dados? Muitas empresas enfrentam esse desafio de frente, buscando novas soluções que alavancam a inteligência artificial (IA), bem como múltiplas capacidades e opções de implantação em nuvens locais, públicas e privadas para inovar sua infraestrutura de dados e negócios.

Este eBook apresenta oito maneiras pelas quais os arquitetos corporativos e CIOs podem modernizar seu gerenciamento de dados para superar os desafios da arquitetura de informações e identificar novas oportunidades de negócios, com opiniões incluídas de analistas do setor e experiências reais de clientes.

Um

Impulsione a otimização e o desenvolvimento das aplicações com a IA

Para serem competitivas, as organizações precisam extrair os insights mais profundos, precisos e acionáveis de seus dados, com rapidez suficiente para implementá-los quando forem mais valiosos. Os insights direcionados aos dados podem estimular a inovação em toda a organização, revelando oportunidades para novos produtos ou mercados, empoderando o pessoal de vendas para ter discussões mais significativas e identificando processos internos que podem ser aprimorados. [Por esse motivo, muitas organizações estão recorrendo às soluções de gerenciamento de dados desenvolvidas e criadas para IA.](#)

Desenvolvido por IA

As soluções baseadas em IA e machine learning otimizam diretamente a performance do gerenciamento de dados, que, por sua vez, tem um efeito cascata em toda a empresa.

Por exemplo, o machine learning pode ser usado para acelerar significativamente os tempos de consulta, melhorando continuamente o caminho que a consulta percorre até os dados. Além disso, ele pode ser usado para classificar os resultados da consulta com base na confiança, para que a obtenção dos melhores insights seja muito mais fácil e eficiente. A incorporação de consultas em linguagem natural também é útil para que os usuários de LOB possam empregar uma interface semelhante à pesquisa na Internet para extrair os insights diretamente dos dados – ajudando a disseminar os insights por toda a organização, removendo gargalos internos e economizando o valioso tempo dos cientistas e analistas de negócios.



Criado para IA

As soluções criadas para IA dão suporte às estruturas, linguagens e ferramentas necessárias para os cientistas de dados e desenvolvedores de aplicações implementarem rapidamente a IA em seus projetos, sem os atrasos das reescritas ao colocar protótipos em produção.

Uma solução criada para IA inclui suporte para linguagens populares como Python, GO e JSON, além dos Notebooks Jupyter. As bibliotecas de ML e ferramentas adicionais permitem que os usuários integrem as aplicações de IA às suas soluções de gerenciamento de dados. Com isso, cientistas de dados e desenvolvedores de aplicações não precisam gastar tempo para aprender e criar um novo código. Em vez disso, eles podem começar a trabalhar nos projetos imediatamente e empregar códigos pré-existentes de trabalhos ou bibliotecas anteriores para concluir o projeto mais rapidamente do que se tivessem começado do zero.

Check-in de gerenciamento de dados

- Seu ambiente de dados oferece suporte às ferramentas e linguagens de ciência de dados mais populares?
- Suas consultas podem ser otimizadas para machine learning e criadas usando a linguagem natural?
- Qual a rapidez com que suas consultas são retornadas? Elas são baseadas na confiança em sua precisão?

Analista anuncia o aumento do gerenciamento de dados com IA

Os analistas estão percebendo a capacidade do gerenciamento de dados de IA em transformar uma empresa. O relatório recente da 451 Research analisa por que 88% dos entrevistados das empresas mais direcionadas a dados disseram: “IA e ML são componentes importantes das plataformas de dados e iniciativas de análise”.

Eles também descrevem toda a variedade de casos de uso de IA, da automação das tarefas repetitivas até as complexas e o aprimoramento das pessoas para que façam todos os processos intermediários.



“IA e ML são componentes importantes das plataformas de dados e iniciativas de análise.”

451 Research: Accelerating AI with Data Management; Accelerating Data Management with AI. [Leia o relatório](#) →

88%

Dois

Permita relatórios e análises rápidos

Diversas tecnologias avançadas estão disponíveis para ajudar a acelerar relatórios e análises, desde computação e processamento até compactação e data skipping. Quando combinadas, elas têm a capacidade de fazer melhorias significativas no ritmo em que os insights podem ser fornecidos. Essas tecnologias são particularmente importantes para os cientistas de dados e analistas de negócios – ajudando a garantir que as consultas sejam retornadas mais rapidamente, permitindo que as empresas tomem decisões em tempo hábil.

Computação em memória com tabelas de sombra baseadas em colunas

A computação em memória se tornou popular devido à sua capacidade de aumentar a performance. No entanto, nem todos os bancos de dados em memória tratam os dados da mesma maneira. Como os bancos de dados precisam da funcionalidade de leitura e gravação, eles são tradicionalmente organizados em linhas e exigem que todos os dados sejam armazenados em memória para atingir as melhorias desejadas na performance. No entanto, com os dados crescendo constantemente, pode não ser possível que todos eles caibam em memória. É por isso que são necessárias tabelas de sombra baseadas em colunas que podem ser analisadas paralelamente às tabelas baseadas em linhas. A leitura de colunas é mais eficiente e as colunas exatas exigidas podem ser armazenadas em memória para análise. Além disso, a movimentação otimizada dos dados de armazenamento para a memória do sistema e para a memória da CPU, encontrada em tecnologias colunares em memória como o IBM® Db2® BLU Acceleration®, acelera o ritmo em que as consultas podem ser concluídas. As cargas de trabalho OLTP e OLAP podem, portanto, ser executadas em paralelo no mesmo banco de dados, o que significa que as consultas podem ser executadas sem a migração para um data warehouse e as respostas podem ser retornadas mais cedo.



Compactação ativa

A compactação ativa permite que os dados sejam digitalizados ou comparados sem descompactá-los. Portanto, a descompactação pode ser limitada aos dados exigidos e adiada até que seja absolutamente necessária. Os bancos de dados conseguem isso compactando cada coluna com seus próprios dicionários de compactação e compactando os valores de dados que aparecem com mais frequência do que outros valores. Isso ajuda a reduzir o tempo até o insight, melhorando a taxa de transferência de E/S, além de economizar energia de processamento e reduzir diretamente os custos gerais de armazenamento.

Data skipping

O data skipping ajuda a economizar tempo durante as consultas, identificando os dados irrelevantes e ignorando-os. Os bancos de dados conseguem isso acompanhando os valores mínimo e máximo das linhas, usando objetos de metadados chamados de tabelas de sinopse. Ao executar consultas, as tecnologias de data skipping analisam as tabelas de sinopse quanto aos intervalos de dados que contêm um valor que se aplica à consulta e ignoram os dados que não correspondem. Apenas os dados exigidos são lidos e a varredura usada é mais eficiente, reduzindo a E/S desnecessária e agilizando a execução da consulta.

Check-in de gerenciamento de dados

- Quanto tempo você gasta esperando a descompactação dos dados durante a análise?
- As consultas são concluídas com rapidez suficiente para você tirar o máximo proveito dos insights?
- Você consegue analisar dados transacionais com eficiência sem interromper as operações?

Como um fabricante global reduziu o tempo de geração de relatórios em 98% usando algumas tecnologias-chave

Para a Knorr-Bremse, fabricante de sistemas de freios, aumentar sua presença global por meio de aquisições e joint ventures significou um grande aumento no volume de dados. Eles precisavam processar até 4 TB de dados por dia. A seleção de tecnologias como compactação ativa, data skipping e computação em memória proporcionou benefícios consideráveis.



“Com o Db2 BLU Acceleration, conseguimos melhorar a performance nos relatórios de 3 minutos para 3 segundos.”

– Thomas Brauchle, diretor de arquitetura e estratégia de TI, Knorr-Bremse

98%

Três

Reduza seu custo total de propriedade

À medida que os orçamentos de TI continuam a crescer, é mais importante do que nunca que as organizações usem os recursos e o pessoal financeiros com eficiência. Os bancos de dados devem fornecer capacidades de gerenciamento de dados líderes do setor e econômicas, além de cumprir os acordos de nível de serviço (SLAs) com uma variedade de serviços totalmente gerenciados e suporte de nível empresarial.

Recursos autônômicos

A automação de tarefas administrativas ajuda nesses dois objetivos, liberando a equipe para se concentrar em iniciativas mais estratégicas e acelerando os fluxos de trabalho. As soluções de banco de dados que podem automatizar uma variedade de tarefas administrativas – como configuração e implantação, gerenciamento de carga de trabalho, utilização de recursos e gerenciamento de armazenamento, bem como manutenção, atualizações e expansão de capacidade – são extremamente valiosas.

É possível economizar tempo e esforço administrativos que poderiam ter sido gastos em reescrita, escolhendo soluções de dados com opções na nuvem e no local que compartilham a base do código nas implantações (mesmo nas nuvens concorrentes). Um bom exemplo disso é a família Db2 de produtos de gerenciamento de dados.



Redução na pegada do armazenamento

O estudo “2019 State of IT” da Spiceworks revelou o que as empresas com mais de 5.000 pessoas dedicaram ao armazenamento: 8% ou mais de seus orçamentos de TI nas categorias de hardware, serviços hospedados e em nuvem e serviços gerenciados. A redução dos requisitos de armazenamento, portanto, pode ter um impacto significativo. Isso pode ser atingido com as tecnologias de compactação de dados e gerenciamento de dados de multitemperaturas, que reduzem diretamente os requisitos de armazenamento ou transferem dados “frios” para opções de armazenamento mais econômicas.

Aqueles que desejam trocar de banco de dados para economizar dinheiro e acelerar processos, sem precisar modificar as aplicações atuais, também podem pensar em procurar soluções com transparência na topologia de cluster. Isso significa que a solução não exige que as aplicações estejam cientes da implantação subjacente de cluster e banco de dados e, portanto, ajuda a acelerar a codificação e o teste, além de tornar os desenvolvedores de aplicações mais produtivos.

Check-in de gerenciamento de dados

- Qual nível de automação está disponível em seu sistema existente?
- Quanto tempo é gasto em reescrita na sua organização?
- Você comparou os preços de gerenciamento de dados entre os fornecedores?

Como a Owens-Illinois observou reduções de custos de sete dígitos após a modernização

Para a fabricante global Owens-Illinois, obter reduções de custos de sete dígitos não é uma tarefa fácil. Quando essas economias também são combinadas com transações mais rápidas e redução da presença de banco de dados e armazenamento próxima de 50%, é uma razão para contar ao mundo. Descubra a história completa no vídeo abaixo.

“Vimos reduções de sete dígitos no custo total de propriedade, 50% de redução no armazenamento e pelo menos 20 a 30% de melhorias na performance no tempo de resposta transacional.”

– Rod Masney, CIO da Owens-Illinois



50%

Quatro

Obtenha flexibilidade para escolher no local ou na nuvem, conforme necessário

Dependendo de suas necessidades exclusivas, considere quais implantações melhor atendem aos seus requisitos: local, nuvem privada, nuvem hospedada, nuvem pública ou um modelo híbrido que utiliza diversas plataformas. Como nem todos os dados são iguais ou têm os mesmos usos, é necessário ter acesso a uma variedade de opções de implantação. A escolha de um fornecedor que ofereça uma variedade de opções de implantação, capazes de oferecer suporte a todos

os aspectos da estratégia da sua organização, fornecerá melhor flexibilidade do que qualquer fornecedor de implantação única. Além disso, se todas as implantações vierem da mesma família de produtos de gerenciamento de dados, é provável que tenham melhor integração do que uma arquitetura montada a partir de soluções pontuais.

Local

As implantações locais fornecem um alto nível de controle sobre otimização, acesso aos dados e visibilidade, mas requerem grandes investimentos em hardware e talento de TI. Elas são usadas com mais frequência quando a segurança é fundamental ou quando são exigidas pelos regulamentos.

Nuvem privada

As implantações em nuvem privada permitem aproveitar a segurança do seu próprio firewall, fornecendo a flexibilidade tradicionalmente associada à nuvem pública. A capacidade de usar a tecnologia de contêiner é um benefício considerável para esse tipo de implantação. Em comparação às máquinas virtualizadas, os contêineres são mais fáceis de implantar rapidamente e geralmente exigem menos espaço de servidor para operar.

Nuvem hospedada

Os ambientes de nuvem hospedada eliminam a responsabilidade de adquirir e manter o hardware, enquanto permitem um certo grau de gerenciamento do sistema. As implantações de nuvem hospedada são ideais para projetos de recuperação de desastre porque você não precisa comprar infraestrutura adicional e tem a possibilidade de usar as licenças de banco de dados existentes. A nuvem hospedada integra-se perfeitamente aos sistemas locais.

Nuvem pública

As implantações em nuvem pública fornecem o menor grau de controle direto sobre aspectos como largura de banda e monitoramento, mas têm flexibilidade aumentada. A adição rápida de capacidade de nuvem pública para complementar as nuvens privadas, conhecida como cloud bursting, separando as necessidades de computação e armazenamento e pagando apenas pelo que você usa, significa que as implantações de nuvem gerenciadas são ótimas opções para backups e ambientes de desenvolvimento ou teste.

Check-in de gerenciamento de dados

- Você pode oferecer suporte a implantações locais e na nuvem na mesma base de código?
- Qual a facilidade de adicionar e integrar diferentes implantações de dados atualmente?
- As opções multinuvem estão disponíveis? Elas estão bem integradas?

Por que um ambiente híbrido holístico e conectado é importante



As empresas não se sentem mais vinculadas à falsa dicotomia da nuvem OU local e estão optando por procurar fornecedores que possam fornecer opções locais E de nuvem bem integradas. Saiba mais sobre as capacidades exigidas para fechar a lacuna entre as diferentes implantações de gerenciamento de dados.

Ponte para a nuvem. [Leia o reporte técnico →](#)



Cinco

Aumente a escalabilidade e a disponibilidade constantes

Manter uma empresa em operação requer sistemas de dados em funcionamento o tempo todo e fortes o suficiente para lidar com a crescente quantidade de dados gerados por uma empresa. O tempo de atividade contínuo e as respostas instantâneas são esperados não apenas pelos clientes e parceiros, mas também internamente. As falhas levam a oportunidades perdidas para gerar negócios adicionais com base em insights, relacionamentos fracos com parceiros e perda de boa vontade. As três opções mais úteis para escalonamento e disponibilidade são descritas aqui.

Armazenamento em cluster compartilhado

Uma opção é o armazenamento em cluster compartilhado, que permite que o trabalho seja distribuído entre nós diferentes que dependem da mesma fonte de dados. Duas principais capacidades para o armazenamento em cluster compartilhado são o balanceamento de carga de trabalho automático entre os nós em sua arquitetura e o failover automático integrado para usar cada nó no armazenamento em cluster compartilhado como um nó de failover para outro. Juntas, essas capacidades aumentam a redundância e evitam sobrecarregar determinados nós para diminuir as chances de falha e fornecer melhor disponibilidade mesmo em situações de desastre.



Suplementação da nuvem local e privada com a nuvem pública

As implantações na nuvem local e privada podem melhorar a escalabilidade e a disponibilidade com a adição da nuvem pública. Para isso, as ofertas de nuvem local e privada são usadas como capacidade fundamental, enquanto as ofertas de nuvem pública são usadas quando armazenamento ou capacidade de computação adicional são exigidos. Uma forte integração entre os sistemas na nuvem e no local é necessária para obter isso eficientemente, para que os dados não fiquem em silos com base na implantação. A ativação de um ambiente de nuvem pública complementar a uma oferta existente de nuvem local ou privada também pode fornecer melhor disponibilidade em caso de interrupções planejadas ou não planejadas na empresa.

Escalonamento independente do armazenamento na nuvem e computação

As empresas devem ser capazes de aumentar ou diminuir independentemente o armazenamento e calcular as ofertas da nuvem para lidar com a variabilidade nas cargas de trabalho específicas da nuvem. Com o escalonamento independente, a nuvem pode aumentar a computação para atender às necessidades de pico de performance e diminuir para economizar fundos quando essa energia adicional não for mais exigida. Maior disponibilidade também exige menos esforço com serviços em nuvem totalmente gerenciados, como recuperação de desastre de alta disponibilidade (HADR), backups diários, manutenção e suporte corporativo.

Check-in de gerenciamento de dados

- Quanto você gasta atualmente em taxas de transferência de dados?
- Suas ofertas na nuvem e no local compartilham uma base de código comum, permitindo uma melhor integração?
- Você consegue escalar o armazenamento do banco de dados na nuvem e computar de maneira independente?

Por que o centro holandês de inteligência de negócios para serviços de saúde usa o escalonamento da nuvem

A Vektis, o centro holandês de inteligência de negócios para serviços de saúde, precisava criar uma versão da verdade, passando do fornecimento de dados em si – que poderia ser alterado e manipulado – para o fornecimento do acesso a uma única versão dos dados que eles coletam. No entanto, a empresa também sabia que precisaria escalar para acomodar múltiplos usuários, todos inserindo consultas difíceis ao mesmo tempo.



“Queríamos migrar para o IBM Db2 Warehouse on Cloud devido à flexibilidade de aumentar e diminuir a escala. Isso nos torna mais ágeis, mas também nos dá o controle dos custos.”

– Herman Bennema, CEO da Vektis

Seis

Ofereça suporte a novos tipos e fontes de dados

Múltiplos tipos de dados e fontes são agora mais prevalentes. Pense no crescimento dramático de dados semiestruturados e não estruturados, gerados a partir de fontes como mídias sociais, sensores em máquinas conectadas à Internet das Coisas (IoT) e repositórios de conteúdo multimídia. Além disso, muitas dessas fontes agora estão produzindo dados de streaming em alta velocidade e volume.

Reunir os dados externos de uma organização com os que a empresa já coleta pode criar insights mais profundos e precisos, capazes de gerar melhores decisões. Por exemplo, usando apenas os dados internos, você poderá ver vendas iniciais robustas para um novo produto. A combinação desses dados com as mídias sociais e outros dados não estruturados externos pode contar uma história diferente, no entanto. Usando a análise de sentimentos, uma empresa seria capaz de dizer se esse produto foi comprado com frequência, mas forneceu experiências negativas – permitindo capturar e corrigir o problema antes da queda nas vendas.

No entanto, a combinação de dados internos e externos apresenta um novo conjunto de desafios de implantação e desenvolvimento para a infraestrutura de dados, que precisa fornecer flexibilidade para atender, capturar, armazenar e processar uma ampla variedade de tipos de informações de diferentes fontes em diferentes velocidades. As empresas precisam de um ambiente de dados que possa servir os dados quando e conforme são necessários e capturar exatamente o que está acontecendo a qualquer momento.

Alguns dos exemplos mais importantes de dados não tradicionais:



O **conteúdo audiovisual** fornece a oportunidade para as organizações aproveitarem conhecimentos que, de outra forma, seriam difíceis de encontrar ou analisar, adicionando informações contextuais valiosas.



Os **dados do evento** fornecem informações que geralmente são muito sensíveis ao tempo. Quando capturados e analisados antes de perder sua relevância, os dados do evento podem orientar as ações em tempo real.



Os **dados de mídia social** podem informar uma visão 360° dos clientes, ajudando a direcionar o marketing, aumentar as vendas e identificar e diminuir rapidamente quaisquer problemas com os clientes ou a percepção geral do público.



Os **dados em documentos físicos** podem ser mais difíceis de acessar, mas permitem que as empresas obtenham insights das informações registradas manualmente, que podem ser mais ricas e ter um contexto mais detalhado do que os formulários on-line comparáveis.



Os **dados do sensor de IoT** podem fornecer alertas quando máquinas ou linhas de produção correrem risco de funcionar mal, incentivando medidas preventivas em oposição a ações reativas.



O **conteúdo de e-mail** pode incluir e frequentemente inclui componentes audiovisuais e também fornece informações mais robustas do que as relatadas, como dados de pesquisas, ou que podem ser observadas, como transações.

Check-in de gerenciamento de dados

- Quantos dados não estruturados você pode acessar e analisar atualmente?
- Com que frequência você combina dados internos e externos para obter insights?
- Quantos dispositivos de IoT sua empresa ou seus clientes usam no dia a dia?

Como a Capitalogix combina dados estruturados e não estruturados para criar um significado contextual para os fundos de hedge

Com as ações mantidas por uma média de 20 segundos, a empresa de tecnologia financeira Capitalogix está em uma busca constante por maneiras de obter insights mais robustos rapidamente e permanecer à frente das tendências do mercado. Para fazer isso, eles usam uma variedade de dados estruturados e não estruturados em diversas velocidades.

“Temos acesso a dados alternativos que podem ser totalmente não estruturados: dados de sentimentos de feeds do Twitter e imagens de satélite.”

– Howard Getson, CEO da Capitalogix



Sete

Integre a tecnologia de código aberto

A tecnologia de código aberto é uma parte fundamental do cenário moderno de gerenciamento de dados. É impossível exagerar na importância da capacidade de aproveitar o esforço de toda a comunidade de desenvolvimento, evitar a dependência de fornecedores e obter as vantagens de mais fontes de dados com opções de código aberto. Quando o código aberto é implementado corretamente, pode levar a custos mais baixos e melhores insights.

Aproveite a inovação da comunidade e evite a dependência de fornecedores

As soluções de código aberto são continuamente aprimoradas por uma comunidade de indivíduos e, às vezes, empresas. O amplo conhecimento técnico e o número de trabalhadores qualificados que podem solucionar qualquer problema ou nova ideia ultrapassam o que uma única empresa pode fazer, levando a tecnologias mais inovadoras. Além dessas tecnologias de ponta, as soluções de código aberto ajudam a reduzir o risco de dependência do fornecedor de diversas maneiras. Como as mesmas tecnologias de código aberto podem ser acessadas de múltiplos fornecedores, é mais fácil selecionar um fornecedor diferente sem o incômodo de recodificar. As habilidades de codificação de código aberto também são transferíveis, diferentemente do conhecimento próprio necessário para os sistemas próprios. Remover a dependência do fornecedor da equação reduz os custos, aumentando as oportunidades de concorrência.



Obtenha acesso a mais dados

As soluções de código aberto também fornecem acesso a mais dados do que aqueles que poderiam ser colocados em bancos de dados tradicionais ou data warehouses. Por exemplo, o Hadoop geralmente é o repositório principal de informações não estruturadas devido à sua capacidade de evitar a atribuição de um esquema ao armazenar dados. Também é possível descarregar dados “frios” no Hadoop a partir de um data warehouse.

Porém, o acesso a mais dados por meio da tecnologia de código aberto vai além do Hadoop. Bancos de dados como MongoDB e PostgreSQL aumentam bancos de dados próprios. O MongoDB, por exemplo, é um banco de dados de documentos baseado em NoSQL. Ao armazenar dados em documentos JSON, ele permite uma rápida evolução do esquema, normalmente não encontrada em bancos de dados relacionais mais estruturados. Isso o torna adequado para casos de uso que envolvem aplicações móveis, análise em tempo real e gerenciamento de conteúdo, todos os quais se beneficiam da flexibilidade do esquema. Da mesma forma, embora o PostgreSQL seja um banco de dados relacional, ele aumenta (em vez de entrar em conflito com) os bancos de dados relacionais existentes em uma organização. Isso porque o PostgreSQL é um banco de dados relacional de objetos, o que significa que ele oferece suporte a objetos definidos pelo usuário e como eles se comportam para criar, armazenar e recuperar estruturas de dados mais complexas.

O PostgreSQL oferece suporte a tipos de dados como UUID, monetários, enumerados, geométricos, binários, endereços de rede, cadeias de bits, pesquisas de texto, XML, JSON, arrays, tipos de composição e intervalo, além de alguns tipos internos para identificação de objetos e localização de log. Embora todo o intervalo de flexibilidade nem sempre seja necessário, ele atua como um complemento útil para os bancos de dados relacionais padrão. Por fim, essas tecnologias suplementares de código aberto permitem o acesso a mais tipos de dados de mais fontes, fornecendo as informações extras necessárias para obter insights mais profundos, capazes de impulsionar a inovação e a otimização dos negócios para obter maiores ganhos financeiros.

Analistas reconhecem o valor total da propriedade da parceria estratégica entre IBM e Cloudera

Os analistas da Cabot Partners entendem que o crescimento e o impacto positivo da inteligência artificial (IA) só são possíveis por meio da arquitetura da informação (AI) correta. Isso inclui soluções de código aberto de nível empresarial, como as fornecidas pela IBM e pela Cloudera. Depois de realizar um estudo sobre Valor total de propriedade (TVO), eles concluíram que “juntas, a IBM e a Cloudera fornecem flexibilidade, escolha e valor sem precedentes para os clientes em sua jornada para industrializar a análise/IA”. Leia o relatório completo para ver como a combinação desses dois líderes de código aberto pode ajudar a fornecer menor custo total de propriedade, produtividade aprimorada, aumento de receita e redução de riscos.



“Juntas, a IBM e a Cloudera fornecem flexibilidade, escolha e valor sem precedentes para os clientes em sua jornada para industrializar a análise/IA.”

Maior escolha e valor para análises avançadas e IA. [Leia o relatório →](#)

Seleção do código aberto correto

As empresas devem garantir que as soluções de código aberto sejam seguras, bem integradas e adequadamente governadas. Embora isso possa ser feito internamente, contar com o suporte de um fornecedor de nível empresarial e as soluções de valor agregado geralmente é mais eficiente. Pense na segurança, por exemplo; soluções corporativas pré-empacotadas com segurança integrada permitem que os desenvolvedores se concentrem em atividades mais valiosas, em vez de dedicar tempo para estabelecer controles de acesso e integrar o sistema atrás do firewall da empresa.

Para integração, a federação de dados no Hadoop e o armazém de dados usando um mecanismo SQL comum é outra área em que as soluções empresariais de código aberto fornecem um valor considerável. Sem essa integração, por exemplo, combinar dados de vendas da primeira semana com dados de mídia social para fornecer uma previsão de demanda futura exigiria muito mais esforço. A governança em nível corporativo também ajuda a reduzir o esforço exigido para reunir dados. Sem os metadados, a catalogação e a organização geral das soluções de código aberto de nível empresarial que compreendem um data lake, pode rapidamente tornar-se difícil de extrair os dados exatos do que você precisa. Seu data lake se torna o que alguns chamam de data swamp.

Check-in de gerenciamento de dados

- Você está atualmente preso a soluções próprias que podem ser substituídas por outras de código aberto?
- Sua empresa possui casos de uso específicos mais adequados para capacidades de banco de dados de código aberto?
- Suas soluções de código aberto atuais têm a segurança, integração e governança necessárias?



Oito

Incorpore novas capacidades de streaming e fast data

Fatores como o incrível crescimento de dados, a crescente prevalência de projetos de IA e uma maior compreensão dos benefícios do código aberto estão mudando a maneira pela qual as empresas capturam e usam dados de streaming. Na verdade, essa mudança é tão proeminente que alguns analistas começaram a se referir a essa categoria de gerenciamento de dados com o termo mais amplo “fast data” em vez de “análise de streaming”, que é mais estreito. A atualização para soluções de fast data pode ajudar as organizações a ingerir e usar os dados rapidamente, incorporar melhor as tecnologias de código aberto e executar análises e IA mais robustas.

Obtenção de velocidades mais altas, mais dados e melhores insights

IoT, fluxo de cliques, aplicações móveis e outras formas de dados que fluem rapidamente a partir de eventos discretos são, por natureza, transitórios. Sendo assim, para tirar o máximo proveito desses valiosos pontos de dados, um sistema de ingestão rápida é crucial. Agora é possível armazenar e analisar até 250 bilhões de eventos por dia com apenas três nós em soluções como o IBM Db2 Event Store. As tecnologias anteriores podem exigir quase 100 nós para realizar o mesmo volume.

Com isso, as organizações se beneficiam de duas maneiras – a principal delas é o custo. A redução nos requisitos de nós ajuda as empresas a eliminar os custos de infraestrutura por meio da eficiência mais alta e a investir esse dinheiro em projetos mais inovadores e significativos. Em segundo lugar, como cada nó é mais eficaz, manter um número semelhante de nós também permite que uma quantidade maior de dados seja capturada para insights mais robustos e completos. Pense nos dados do fluxo de cliques; cada evento adicional capturado é uma oportunidade de entender e responder melhor a um cliente enquanto ele está em um site, fornecendo a capacidade de se adaptar às suas necessidades e, com sorte, aumentar os lucros.



Reforço dos fast data com o código aberto

Muitos dos mesmos benefícios do código aberto discutidos na seção anterior – como redução da dependência do fornecedor e acesso a mais dados – também se aplicam quando o código aberto é incorporado em soluções de fast data. No entanto, a velocidade dos eventos e a transmissão de dados impõem desafios únicos. As soluções que podem gravar em um armazenamento de objeto de formato aberto, como o Apache Parquet, também se tornaram populares recentemente para armazenar grandes quantidades de dados de streaming, porque podem ser mais rápidas e custar menos do que outros ambientes como o Hadoop. A capacidade de capturar dados de eventos em velocidades incrivelmente altas depende tanto do repositório quanto do mecanismo de análise e ingestão.

Realização de análises e IA mais robustas

As soluções mais recentes de fast data levam em consideração a ampla variedade de casos de uso de análise e IA que devem ser abordados, facilitando a combinação de dados históricos e de streaming para obter insights. Como a análise em tempo quase real é importante para atuar em muitos tipos de dados que as soluções de fast data são construídas para capturar, é importante que o consumo de dados não seja feito aleatoriamente. Otimizar a organização por meio da otimização da memória pode ajudar a facilitar o uso dos dados. Um bom exemplo disso é o monitoramento de alertas. Os dispositivos de IoT que monitoram sistemas críticos enviam grandes volumes de dados de streaming de alta velocidade com a finalidade de alertas. A capacidade de analisar dados que foram organizados adequadamente durante a ingestão facilita a ação quase instantaneamente, em vez de precisar esperar até que alguém atribua uma estrutura. Quando alertas críticos estão em jogo, a ação instantânea faz a diferença.

Além disso, a combinação de dados de streaming no momento com dados históricos ou recém transmitidos permite um melhor contexto para a análise. Os dados recém transmitidos mostram se um ponto de dados faz parte de uma tendência ou de uma ocorrência pontual. Os dados históricos podem ser usados para determinar quais pontos de dados geralmente ocorrem antes de uma falha digna de alerta. Essa também é uma ótima oportunidade para introduzir o machine learning para refinar continuamente o que é considerado digno de alerta. Ao fazer isso, um sistema de monitoramento de alertas pode se tornar mais preciso, minimizando o aborrecimento geralmente associado a alarmes falsos e ainda ajudando a evitar falhas críticas.

Check-in de gerenciamento de dados

- Atualmente, você consegue combinar dados históricos e de streaming para decisões quase em tempo real?
- Quantos nós você usa atualmente e quantos eventos eles permitem que você processe?
- Ser capaz de usar um formato de dados aberto para armazenamento de objetos forneceria a flexibilidade necessária?

O estudo recente da Forrester Consulting revela mais sobre as rápidas mudanças no cenário do fast data

A Forrester Consulting realizou recentemente um estudo sobre fast data (incluindo dados de streaming e de eventos) que reafirmou sua importância contínua. A Forrester pesquisou indivíduos para descobrir a prevalência do fast data, tanto em termos de uso atual quanto em planos futuros. Eles também observaram os obstáculos que podem reduzir a eficácia de soluções de fast data e forneceram algumas recomendações importantes.



“88% dizem que precisam executar análises em tempo quase real nos dados transmitidos armazenados.”

Não fique travado esperando pelo fast data.

[Leia o estudo da Forrester Consulting →](#)

88%

Conclusão

Como começar a modernizar seu gerenciamento de dados

A dependência das organizações em relação aos dados está se aprofundando a cada dia. Esse recurso está crescendo rapidamente, com mais tipos e fontes do que nunca, apresentando oportunidades e desafios. É por isso que é crucial ter uma infraestrutura de gerenciamento de dados híbrida que atenda a essas necessidades e cresça com elas.

Para saber mais sobre o IBM Hybrid Data Management, entre em contato com o seu representante IBM ou Parceiro de Negócios IBM. Você também pode [iniciar uma avaliação gratuita do IBM Cloud Pak for Data](#).

Se você tiver alguma dúvida sobre este eBook, o gerenciamento de dados híbridos ou o IBM Cloud Pak for Data, nossos especialistas ficarão felizes em agendar uma conversa gratuita de 30 minutos para te ajudar.

Converse com um especialista →

IBM HDM e Cloud Pak for Data fazem a diferença

Também é crucial que sua arquitetura seja voltada para o futuro, incorporando as ferramentas de IA, ML e ciência de dados necessárias para continuar a se destacar nesse ambiente competitivo, oferecendo opções para implantação no local, na nuvem privada ou em diversos fornecedores de nuvem pública. Também é essencial ter uma ampla variedade de ofertas de gerenciamento de dados, incluindo banco de dados, data warehouse, data lake e capacidades de fast data.

O IBM Hybrid Data Management fornece as capacidades de IA, opções de implementação e variedade de soluções necessárias. O IBM Db2 é desenvolvido e criado para a IA, melhorando a performance e fornecendo as ferramentas necessárias para criar aplicações de IA. Os produtos IBM Db2 do mecanismo SQL comum são baseados em capacidades aprimoradas de integração e capacidade de escrever uma vez e executar em qualquer lugar.

Os produtos IBM Hybrid Data Management estão disponíveis como parte do IBM Cloud Pak for Data. Ao executar soluções em um contêiner Kubernetes com o Cloud Pak for Data, essas soluções se tornam agnósticas para hardware e independentes de nuvem – capazes de executar em qualquer lugar que ofereça suporte a contêineres Linux ou RedHat, respectivamente. Como um benefício adicional, a virtualização de dados está incluída no Cloud Pak for Data, permitindo executar cargas de trabalho com mais eficiência sem mover os dados.

Aumentar sua capacidade também é simples com o Cloud Pak for Data. Basta adicionar VPCs adicionais para acessar recursos adicionais. Isso vai além da extensão do gerenciamento de dados híbridos e das ofertas básicas a soluções como o IBM Watson Studio Open Scale, para que todas as partes do IA Ladder permaneçam conectadas – permitindo que os dados sejam automaticamente controlados e usados sem problemas em projetos de IA.



© Copyright IBM Corporation 2020

IBM Corporation
Route 100
Somers, NY 10589

Produzido nos Estados Unidos da América
Março de 2020

IBM, o logotipo IBM, **ibm.com**, Db2 e BLU Acceleration são marcas comerciais da International Business Machines Corp., registradas em diversas jurisdições no mundo todo. Outros nomes de produtos e serviços podem ser marcas comerciais da IBM ou de outras empresas. Uma lista atual das marcas comerciais da IBM está disponível na web em “Copyright and trademark information” (Informações de copyright e marcas comerciais), no link www.ibm.com/legal/copytrade.shtml.

Este documento encontra-se atualizado na data inicial de sua publicação e pode ser alterado pela IBM a qualquer momento. Nem todas as ofertas estão disponíveis em todos os países em que a IBM opera.

Os exemplos de clientes e dados de performance citados foram apresentados apenas para fins ilustrativos. Os resultados de performance reais podem variar dependendo das configurações específicas e das condições operacionais.

AS INFORMAÇÕES CONTIDAS NESTE DOCUMENTO SÃO FORNECIDAS “NA FORMA EM QUE SE ENCONTRAM” SEM QUALQUER GARANTIA, EXPRESSA OU IMPLÍCITA, INCLUINDO NENHUMA GARANTIA DE COMERCIALIZAÇÃO, ADEQUAÇÃO A UMA DETERMINADA FINALIDADE E NENHUMA GARANTIA OU CONDIÇÃO DE NÃO-VIOLAÇÃO. Os produtos da IBM são garantidos de acordo com os termos e condições dos acordos sob os quais eles são fornecidos.