

WHITE PAPER

Abordando os desafios de armazenamento para suportar cargas de trabalho de análise de negócios e big data

Patrocinado por: IBM

Laura DuBois
Setembro de 2012

RESUMO EXECUTIVO

Atualmente, os projetos de análise de negócios são iniciados para aprimorar operações dos clientes e de negócios em quase todos os setores, exercendo um efeito transformador sobre diversas empresas. As abordagens de data warehousing tradicional e do processamento analítico online, combinado com a execução do novo processamento analítico em relação ao fluxo de dados e dados em tempo real, permitem que as empresas reajam dinamicamente a diferentes clientes, usuários, fornecedores e outras partes interessadas importantes. O que era apenas informação, fornecida periodicamente em um momento específico, se transformou em insight com dados e análises disponíveis em tempo real a qualquer momento.

As vantagens que as empresas podem extrair da análise de negócios e de projetos de big data são inúmeras. O crescimento dos clientes, a identificação dos clientes mais lucrativos e as crescentes taxas de retenção e de medição são benefícios observados pelos provedores de telecomunicações. As centrais de atendimento para empresas de diversos setores podem alavancar os processos analíticos a fim de medir e melhorar a eficiência operacional e/ou o atendimento ao cliente. Instituições bancárias são capazes de atenuar riscos de negócios, segurança e privacidade; reduzir fraudes e gerenciar a conformidade. Além disso, provedores de assistência médica podem transformar e automatizar processos operacionais tanto financeiros quanto de outra natureza, bem como personalizar cuidados aos pacientes.

VISÃO GERAL DA SITUAÇÃO

Big Data e os quatro Vs

A convergência de dispositivos inteligentes (produtos de consumo como, por exemplo, smartphones e tablets, carros, prédios e infraestrutura inteligentes, etc.), redes sociais, redes de banda larga disseminada e usuários de softwares analíticos anunciam uma nova era de análise de negócios que está redefinindo as relações entre produtores, distribuidores e clientes de bens de consumo e serviços.

No passado, as empresas precisavam lidar apenas com um número limitado e gerenciável de fontes de dados. Contudo, o ambiente de negócios atual inclui não somente mais dados, como também mais tipos de dados como nunca incluiu antes. A combinação de dados de uma variedade de fontes e em uma variedade de formatos é um desafio essencial com o qual projetos de análise de negócios e de big data devem lidar. Outro componente de big data é a velocidade, ou a velocidade com a qual as informações são recebidas, analisadas e entregues. A velocidade dos dados se movendo pelos sistemas de uma organização varia desde a integração em lotes e o carregamento de dados em intervalos pré-determinados até o fluxo dos dados em tempo real. O primeiro pode ser visto em data warehousing tradicional e é também atualmente o principal método de processamento de dados que utiliza o sistema Hadoop. Este último é o

domínio de tecnologias como o processamento de eventos complexos (complex event processing ou CEP), mecanismos de regras, busca e análises de texto, inferência, aprendizagem de máquina e arquiteturas baseadas em eventos em geral. Projetos de análise bem-sucedidos requerem a informação correta no momento apropriado com o grau de precisão adequado.

No contexto de big data, o valor se refere ao custo da tecnologia e ao valor derivado do uso de big data. O valor pode ser amplamente observado a partir de uma perspectiva de infraestrutura como também de uma perspectiva de negócios. Os benefícios do negócio podem incluir a eficiência operacional e os aprimoramentos do processo de negócios. Os ganhos em eficiência operacional são medidos pela redução nos custos de mão de obra devido a métodos mais eficientes para a integração, gerenciamento, análise e entrega de dados. Os aprimoramentos do processo de negócios são medidos por um aumento na receita ou lucro devido a novas ou melhores maneiras de condução dos negócios, incluindo melhorias nas transações comerciais, gerenciamento sustentável de comunidades e distribuição apropriada dos serviços sociais, educacionais e de assistência médica. O quarto atributo de big data é o volume. Os projetos de big data tendem a implicar terabytes a petabytes de informação. No entanto, alguns segmentos do mercado e organizações provavelmente possuem poucos gigabytes ou terabytes de dados em comparação aos petabytes ou exabytes de algumas das organizações de redes sociais. Entretanto, esses aplicativos aparentemente menores são capazes de exigir análise e processamento de informações intenso e complexo que caracterizam os aplicativos de big data.

Os desafios de armazenamento com cargas de trabalho de análise de negócios e de big data

Atualmente, as organizações de armazenamento precisam gerenciar o crescimento explosivo da infraestrutura e da capacidade de armazenamento enquanto reduz os custos associados aos crescentes conjuntos de dados. Volumes de dados tendem a dobrar anualmente. No entanto, ao mesmo tempo que os dados primários continuam a aumentar, os orçamentos e o número de recursos de TI para gerenciar sua capacidade crescente permanecem os mesmos. O crescimento nos dados corporativos provém, em parte, do aumento do número de dispositivos corporativos conectados, novos dados de programas e aplicativos sociais, além do desejo de se obter mais informações dinâmicas e em tempo real por toda a corporação.

Além do contínuo crescimento dos dados, as empresas enfrentam exigências legais, reguladoras e de negócios para reter dados de uma variedade de diferentes fontes de conteúdo. A retenção desses dados permite que as empresas preservem sua memória institucional. Além disso, os dados estão disponíveis para gerar valor de negócios no futuro. Cada vez mais, as empresas estão aproveitando dados de conteúdo fixos ou históricos para a análise de dados. Setores como o da saúde aproveitam os dados arquivados nos prontuários de pacientes e outras pesquisas clínicas para estudar as taxas de mortalidade. Os provedores de telecomunicações devem manter os registros de chamadas durante determinados períodos de tempo, porém também estão analisando esses dados para prever o futuro comportamento dos clientes e aprimorar o atendimento ao cliente. Serviços financeiros, provedores de assistência médica e seguradoras aproveitam grandes volumes de dados para detectar e isolar fraudes.

No entanto, equipes de infraestrutura de armazenamento devem não só enfrentar desafios da infraestrutura de crescimento de dados e da retenção de conteúdo fixo por longos períodos de tempo, como também responder às demandas de negócios mais rapidamente. Essas demandas de negócios apresentam-se na forma de novos aplicativos de clientes, novos programas de negócios e novos projetos de análise de negócios – todos exigindo uma infraestrutura de armazenamento escalável, otimizada e resiliente.

OS OBSTÁCULOS NA REALIZAÇÃO DE PROJETOS DE ANÁLISE DE NEGÓCIOS BEM-SUCEDIDOS

As obrigações financeiras de uma análise de negócios bem-sucedida requerem um conjunto amplo e detalhado de habilidades e recursos, incluindo software de análise de negócios e conteúdo, software de integração de informações para conectar informações provenientes de fontes de dados díspares, serviços de análise de negócios e a infraestrutura de armazenamento adequada. Para oferecer suporte à análise de negócios de maneira efetiva, os profissionais de armazenamento e infraestrutura devem continuar a procurar formas mais econômicas e efetivas de armazenar dados ao mesmo tempo que garantem que os objetivos de resiliência e de escalabilidade para cargas de trabalho de análise de negócios não somente sejam alcançados, mas também superados.

Os volumes de informação sobre as soluções de análise de negócios e melhores práticas são frequentemente negligenciados para destacar o impacto da infraestrutura de hardware sobre o sucesso dos projetos de análise de negócios. Acredita-se que os projetos de análise de negócios representam um requisito único, homogêneo que abrange toda a empresa. Esta premissa conduz a diversos conceitos equivocados de mercado que resultam no desempenho deficiente do sistema, arquitetura rígida e manutenção onerosa – em outras palavras, projetos falhos. A realidade é que:

- ☒ Projetos de análise de negócios é um termo de proteção que reúne múltiplas cargas de trabalho relacionadas, requisitos de suporte e automação da decisão do usuário final e tecnologias de computação e armazenamento de alto desempenho.
- ☒ Uma combinação de infraestrutura de servidor e armazenamento que seja expansível e escalável pode ser capaz de suportar coleta e análise de dados. Nenhuma abordagem irá abranger todos os casos de utilização.
- ☒ Clientes corporativos devem considerar como melhor suportar fluxos de trabalho complexos com uma gama de implementações de servidores e armazenamentos. O posicionamento dos recursos de TI frequentemente influencia a latência e o desempenho da carga de trabalho geral de ponta a ponta. Além disso, a infraestrutura da rede possui um suporte robusto sobre o resultado das tecnologias baseadas em servidor e em armazenamento que suportam cargas de trabalho analíticas.

A importância das considerações sobre infraestrutura para a implementação de análise de negócios bem-sucedida

A ausência de uma estratégia de infraestrutura efetiva de armazenamento para cargas de trabalho de análise de negócios pode frequentemente resultar em problemas de desempenho, custos inesperados e insatisfação da unidade de negócio.

Quando o “conjunto” de dados era primeiramente um processamento de transações online (online transaction processing ou OLTP), a geração de 1s e 0s era o foco principal de qualquer projeto de TI, e impulsionar o desempenho era o caminho mais importante à inovação para um projeto.

Atualmente, nenhum indivíduo ou departamento de negócios pode absorver e analisar todos os dados gerados. Na verdade, múltiplas fontes de dados estão oferecendo às organizações de TI algo para pensar: dados de engenharia, de assistência médica, de transporte/logística e, com destaque, dados de mídia social gerados por websites e telefones celulares. Portanto, novas abordagens devem ser desenvolvidas para reunir dados multiestruturados como também armazená-los e analisá-los em tempo hábil.

IBM SMARTER STORAGE PARA CARGAS DE TRABALHO DE ANÁLISE DE NEGÓCIOS

A IBM está acelerando a iniciativa Smarter Computing ao aumentar seus recursos de escalabilidade, otimização e resiliência de suas soluções de armazenamento que, em conjunto com sistemas de computação técnica da IBM, são o fundamento da análise de negócios. A IBM possui uma abordagem estratégica para projetar e gerenciar a infraestrutura de armazenamento para uma maior automação e inteligência. Essas ofertas ajudam os clientes corporativos a obterem resultados analíticos mais rapidamente e a atenderem aos objetivos de crescimento enquanto oferecem mais economia para as cargas de trabalho da análise de negócios.

No entanto, a avaliação e a aquisição da infraestrutura de análise de negócios dependem de diversas variáveis. A tabela 1 destaca algumas das variáveis que uma organização enfrenta e como elas são mapeadas para que as ofertas IBM Smarter Storage otimizem as cargas de trabalho de análise de negócios.

TABELA 1

Considerações sobre infraestrutura e carga de trabalho analítica

Considerações sobre analítica	Considerações sobre negócios	O aspecto do projeto de infraestrutura	Recursos do IBM Smarter Storage
Considerações sobre carga de trabalho analítica			
OLAP (Online Analytical Processing)	Velocidade da saída de consulta versus flexibilidade ad hoc	Velocidade para construir cubos ou metadados de cubos	Empilhamento, uso de SSDs (Solid State Drives), armazenamento de alto desempenho.
Análise detalhada	Capacidade de considerar todas as informações necessárias para um insight detalhado.	Escala e complexidade; manipulação de dados da classe de petabytes, uniões complexas, leitura pesada	Virtualização do armazenamento, armazenamento de expansão e escalabilidade
Análise operacional	Velocidade do insight necessário para os processos de negócios, especialmente para atendimento ao cliente	Simultaneidade de usuários e latência de acesso de dados e computação	Empilhamento, utilização de SSDs, armazenamento próximo à camada de computação.
Considerações características da carga de trabalho			
Variedade, velocidade e volume de dados	Capacidade de confiar em dados para tomada de decisão implícita	Alimentação rápida e confiável de dados; gerenciamento de armazenamento	Gerenciamento de armazenamento, armazenamento unificado, compactação, thin provisioning e outros recursos de eficiência
Variedade de análise, integração de modelos, análise e saída do modelo	Velocidade de resultados versus precisão	Comunicação do interprocessador, largura de banda da rede	Empilhamento, utilização de SSDs, armazenamento próximo à camada de computação.

TABELA 1**Considerações sobre infraestrutura e carga de trabalho analítica**

Considerações sobre analítica	Considerações sobre negócios	O aspecto do projeto de infraestrutura	Recursos do IBM Smarter Storage
Considerações sobre implementação organizacional			
Número de usuários e método de acesso	QoS (Quality of Service), SLA (Service-Level Agreement) e capacidade de agir no momento do impacto	Simultaneidade, largura de banda da rede	Auto-otimização de posicionamento de dados
Análise interativa ou "push" de informações	Análise dinâmica em tempo real versus análise estática	Gerenciamento de recursos, rendimento de E/S, fornecimento	Gerenciamento de armazenamento, virtualização de armazenamento, empilhamento, uso de SSDs
Insights fora da empresa com clientes e parceiros	Eficiência da cadeia de valores, satisfação do cliente	Segurança e fornecimento	Virtualização do armazenamento

Fonte: IDC, 2012

Para um engajamento de análise bem-sucedida, as empresas devem reconhecer que a análise de negócios está intimamente vinculada à infraestrutura de armazenamento. Para obter o máximo de sucesso, as empresas devem criar um sistema de informações escalável, eficiente e confiável, além de uma base de armazenamento que melhore os aspectos econômicos da TI e otimize o desempenho da carga de trabalho da análise. A infraestrutura de armazenamento deve ser capaz de suportar e otimizar as cargas de trabalho que satisfaçam a tomada de decisões complexas, identificando tendências e outliers e prevendo resultados utilizando tecnologias paralelas de alto desempenho. Além disso, arquiteturas resilientes são importantes no suporte à análise em escala, aos sistemas de missão-crítica confiáveis que lidam com um grande número de usuários de forma segura e discreta.

Escalável

Atualmente, diferentes tipos de análises, incluindo OLAP (Online Analytical Processing), data warehousing, fluxo de dados, série temporal e análise detalhada, necessitam de diferentes recursos de computação e armazenamento altamente escaláveis. A criação de uma base de armazenamento escalável e eficiente melhora os aspectos econômicos da TI e otimiza o desempenho da carga de trabalho analítica utilizando todos os dados e informações disponíveis. O IBM Smarter Storage é escalável e eficiente em função do seu projeto, fornecendo os principais recursos necessários para análise mais inteligente, incluindo:

- ☒ **Compactação.** A solução Real-time Compression da IBM pode ser implementada no controlador ou em um aparelho individual e é capaz de compactar dados primários ativos, oferecendo uma redução no custo de até 40% por terabyte. Os dados da carga de trabalho analítica e, em particular, os fluxos de dados, podem ser dinamicamente escalados, e dentro da estrutura de armazenamento atual com baixo custo por gigabyte.

- ☒ **Expansão do armazenamento.** A IBM suporta expansão de arquiteturas de armazenamento de arquivo e blocos que permitem o desempenho de escala horizontal e capacidade como E/S e determinações necessárias ao armazenamento. Escala sem interrupções pode ser realizada enquanto a infraestrutura permanece online e é necessário o envolvimento mínimo das equipes de armazenamento, assim, os processos analíticos não são impactados.
- ☒ **Utilização do armazenamento.** Os recursos de eficiência em armazenamento da IBM, tais como thin provisioning, virtualização de armazenamento e hierarquia de armazenamento podem fornecer a utilização do armazenamento ideal. A utilização do armazenamento pode ser aumentada em até 50%, além de escalar a infraestrutura de armazenamento existente para cargas de trabalho analíticas de alto crescimento.

Otimizado

A análise requer, a qualquer hora e em qualquer lugar, um sistema otimizado para suportar a análise a qualquer momento. Sistemas otimizados são aqueles sistemas sintonizados para alocar os recursos certos no momento certo. O armazenamento também pode ser otimizado para assegurar a melhor utilização do armazenamento ao menor custo. O IBM Smarter Storage é auto-otimizado, fornecendo os principais recursos necessários para a análise de negócios, incluindo:

- ☒ **Posicionamento de dados eficiente.** O suporte às complexas análises em tempo real requer a capacidade de posicionar os dados, de maneira otimizada, na camada correta do armazenamento para atender aos requisitos de desempenho. O IBM Easy Tier, um recurso do DS8000, o Storwize V7000 e o SVC oferecem desempenho três vezes melhor que o IOPS com apenas 3% dos dados em SSDs (Solid State Drives).
- ☒ **Autoajuste.** Projetos de análise dão ênfase à integração e análise de dados. Devido à natureza dinâmica dos projetos de análise, a complexa infraestrutura de armazenamento que requer sobrecarga manual não é desejada. O armazenamento deve ser autogerenciado caso a configuração inicial tenha ocorrido. O armazenamento deve expandir dinamicamente conforme a necessidade e os dados devem ser balanceados em todos os recursos no sistema. O armazenamento IBM inclui diversos recursos de autoajuste como o recurso de distribuição automática de dados do IBM XIV Storage System, que elimina tarefas tradicionais de gerenciamento de armazenamento.

Resiliente

O incentivo à tomada de decisão com base em dados requer uma infraestrutura de TI resiliente e capaz de suportar a proliferação para um grande número de usuários de maneira ininterrupta e segura. Ao implementar a análise, as arquiteturas resilientes podem estar no local ou na nuvem. Tanto a resiliência quanto a virtualização são essenciais para tornar a nuvem ágil, um pilar importante para as soluções IBM Smarter Storage. O IBM Smarter Storage permite que as empresas obtenham resiliência e acessibilidade para suas cargas de trabalho analíticas.

- ☒ **Virtualização do armazenamento.** O IBM SAN Volume Controller e o SmartCloud Virtual Storage Center virtualizam os recursos de armazenamento. O armazenamento IBM também inclui virtualização integrada para facilitar a implementações em nuvem.
- ☒ **Automação do fornecimento.** As empresas podem permitir a análise no ponto de impacto com a automação de IaaS e com o catálogo de serviços de armazenamento, que vincula os requisitos do usuário e os recursos TI.

- ☒ Controles baseados em política. O IBM Active Cloud Engine habilita a fácil criação e aplicação das políticas de arquivo.
- ☒ Proteção, recuperação e retenção. Os recursos de fita, sistemas de backup com base em disco e backup e arquivamento da IBM incluem inovações de fita líder de mercado e oferece suporte para a automação baseada em política.

DESAFIOS/OPORTUNIDADES

As empresas procuram aproveitar o poder da análise quanto à eficiência, inovação ou controle. No entanto, os objetivos organizacionais precisam ser entendidos, os requisitos de usuário devem ser definidos, as fontes e tipos de dados devem ser identificados, o armazenamento correto e a infraestrutura de computação para TI e equipes da unidade de negócios devem ser selecionados, e programas em andamento devem ser estabelecidos para reavaliar continuamente todos os fatores anteriores.

A IBM possui um amplo portfólio de produtos e serviços testados no mercado para atender aos requisitos da análise de negócios. Suas ofertas incluem infraestrutura e software que foram otimizados para suportar cargas de trabalho analíticas. E a IBM irá certamente competir com outras grandes empresas, que veem a mesma oportunidade, e também com uma série de empresas menores, que podem trabalhar para romper o “status quo” e derrubar o negócio tradicional com novas tecnologias e abordagens à análise. A gama de ofertas de análise de negócios da IBM, como o software analítico de negócios e conteúdo, além de integração de informações e infraestrutura de TI para análise de negócios, pode diferenciar a IBM das demais empresas.

RECOMENDAÇÕES

A infraestrutura de armazenamento não pode, e não deve, ser uma consideração tardia. Como os clientes que adotaram a análise de negócios podem atestar, os sistemas analíticos mais flexíveis, escaláveis e resilientes foram desenvolvidos – e colocados em produção – por meio de implementações cuidadosas. As organizações inseridas na economia inteligente devem considerar as melhores práticas a seguir:

- ☒ Desenvolver uma estratégia analítica de negócios (ou analisar uma já existente) que inclua infraestrutura de TI. As empresas devem abordar componentes estratégicos como, por exemplo, tipos e tomadores de decisão, métricas e KPIs, requisitos de latência da informação, fontes e tipos de dados, como também tecnologia e serviços.
- ☒ Reconhecer que uma medida (tecnologia) não se adequa a tudo (requisitos de análise de negócios). Diferentes cargas de trabalho, tipos de dados e tipos de usuários são mais bem atendidos pela tecnologia construída com o propósito de abordar um caso de uso específico. Considerar a infraestrutura de armazenamento para diferentes cargas de trabalho analíticas. Há uma oportunidade de implementar a infraestrutura de armazenamento otimizada para um caso de uso e software específicos.
- ☒ Determinar os requisitos de infraestrutura e armazenamento em paralelo. Apesar de ser provável que os usuários corporativos ofereçam o máximo de seus esforços em relação aos requisitos de software, os grupos de TI devem assegurar que a seleção da infraestrutura de hardware não se tornará uma consideração tardia. Por exemplo, o acesso em tempo real aos dados para realizar uma avaliação rápida do cenário pode justificar uma estratégia de armazenamento diferente do que acessar petabytes de dados. Empresas como Vestas, Bank of America e Walmart atingiram a escala de petabytes.

- ☒ Considerar o impacto do desempenho com a habilitação da infraestrutura de armazenamento. A opção de utilizar a computação em memória, bancos de dados analíticos MPP, aparelhos ou componentes separados terá um impacto material sobre a estratégia de armazenamento bem como a solução de análise de negócios.
- ☒ Avaliar de maneira cuidadosa e testar repetidamente os “insumos e velocidades” da infraestrutura analítica como escalas do projeto – em termos de tamanho da infraestrutura do sistema e da quantidade de dados a serem analisados ou em termos de número de usuários com acesso aos dados. As principais considerações incluem a capacidade dos servidores e dispositivos de armazenamento, quantidade e tamanho dos caches de dados, bem como a transferência entre nós e dentro dos nós associada à latência. A reflexão deve ser direcionada às práticas como, por exemplo, a deduplicação de dados redundantes e a verificação da integridade para otimizar recursos e garantir resultados analíticos válidos.

CONCLUSÃO

Atualmente, as equipes de armazenamento e infraestrutura estão sendo convocadas para não somente gerenciar a infraestrutura atual, mas também oferecer suporte, tanto aos projetos de análise de negócios já existentes quanto aos novos. Os projetos de análise de negócios suportam necessidades de negócios concretos e fornecem informações acessíveis aos tomadores de decisão, incluindo executivos, funcionários da linha de negócios e sistemas automatizados.

Ainda assim, a tecnologia de armazenamento e a infraestrutura de suporte à análise de negócios são imprescindíveis para o sucesso do projeto. A otimização, resiliência e escalabilidade da infraestrutura de armazenamento são capazes de separar o sucesso de problemas por meio de um projeto analítico. Com os grandes volumes de alta velocidade, dados multiestruturados que as empresas devem explorar e analisar, é possível que um projeto de análise de negócios carregue diversos problemas.

As equipes de armazenamento, executivos de TI e usuários corporativos obterão benefícios ao reconhecer que a implementação da infraestrutura de armazenamento adequada, a qual suporta uma ampla gama de cargas de trabalho de análise de negócios, exigirá avaliação constante e boa vontade para ajustar a infraestrutura conforme a necessidade. Isso significa que a flexibilidade do projeto é uma consideração essencial. A capacidade de resposta dos sistemas resultantes é altamente importante para o sucesso dos projetos analíticos. A quantidade de tempo que os usuários finais necessitam para encontrar suas “respostas” aos seus negócios é a chave para o sucesso do projeto e para a percepção dos usuários corporativos sobre a qualidade do desempenho do grupo de TI interno.

A IBM está em uma posição vantajosa por oferecer uma gama de soluções para os atuais projetos de análise de negócios. No entanto, a IBM está dando um passo além na aceleração de sua iniciativa Smarter Computing aprimorando os recursos de escalabilidade, otimização e resiliência de suas soluções de armazenamento que, em conjunto com seus sistemas de computação técnica, oferecem uma base sólida para a análise de negócios atual e futura.

Este documento foi desenvolvido com recursos da IBM. Embora o documento utilize material publicamente disponível de vários fornecedores, incluindo a IBM, ele não reflete necessariamente as opiniões de tais fornecedores sobre as questões abordadas neste documento.

Aviso de Copyright

Publicação Externa de Informação e Dados da IDC — Qualquer informação da IDC a ser utilizada em publicidade, comunicados da imprensa ou materiais promocionais requerem aprovação prévia por escrito do Vice-Presidente ou Gerente de País da IDC. Um rascunho do documento proposto deve acompanhar qualquer pedido relacionado. A IDC se reserva ao direito de negar a aprovação do uso externo não importando o motivo.

Copyright 2012 IDC. A reprodução sem a permissão por escrito é expressamente proibida.