

VALIDAÇÃO ECONÔMICA

Vantagens econômicas do Google Cloud BigQuery em relação às soluções alternativas de EDW baseadas em nuvem


Um EDW realmente baseado em nuvem sem servidor que oferece simplicidade operacional, escalabilidade e flexibilidade inigualáveis e economia de até 54% em relação a EDWs alternativos baseados em nuvem

Por Aviv Kaufmann, diretor de prática e analista principal de validação econômica do Enterprise Strategy Group

Fevereiro de 2024


Conteúdo

Introdução	3
Desafios	3
A solução: plataforma de dados corporativos sem servidor Google Cloud BigQuery	4
Comparação de BigQuery com serviços alternativos de EDW baseados em nuvem	6
Validação econômica do Enterprise Strategy Group	8
Visão geral do valor econômico: comparação do BigQuery com soluções alternativas de EDW baseadas em nuvem	8
Eliminação de investimento inicial e planejamento conforme o porte da implantação	8
Redução nas despesas operacionais	12
Maior agilidade nos negócios e redução no custo da administração diária do EDW	13
Considerações sobre migração	16
Análise do Enterprise Strategy Group	16
Gastos iniciais com nuvem (investimento de capital pago antecipadamente).....	18
Custos de nuvem sob demanda	19
Custos administrativos	19
Questões a considerar	20
Conclusão	21



Validação econômica: resumo das principais descobertas

Benefícios validados do Google Cloud BigQuery



TCO de 28% a 54% menor em comparação com ofertas de EDW de nuvem alternativa

- Nenhum investimento inicial é necessário
- Preços justos e flexíveis
- Projeto realmente sem servidor com a mais alta escalabilidade e simultaneidade
- Armazenamento e escalonamento automático do BigQuery otimizados para desempenho e economia
- Suporte para IA/ML e GenAI

- **Eliminação do investimento inicial e do planejamento:** o preço flexível e previsível de computação e storage elimina a necessidade de pagar meses ou até mesmo anos com antecedência para reduzir os gastos com nuvem e elimina a adivinhação e o planejamento necessários para dimensionar os recursos.
- **Redução nas despesas operacionais:** elimina a necessidade de gerenciar nós virtuais do EDW, bem como a necessidade de monitorar, solucionar problemas, atualizar, ajustar e planejar o crescimento.
- **Maior agilidade nos negócios:** pode ser ampliado ou diminuído, conforme necessário, para atender às demandas de negócios em constante mudança, permitindo que as organizações atuem rapidamente em novas oportunidades sem a necessidade de planejar requisitos de configuração, pausar bancos de dados ou acelerar instâncias de armazém dedicadas para cada organização.

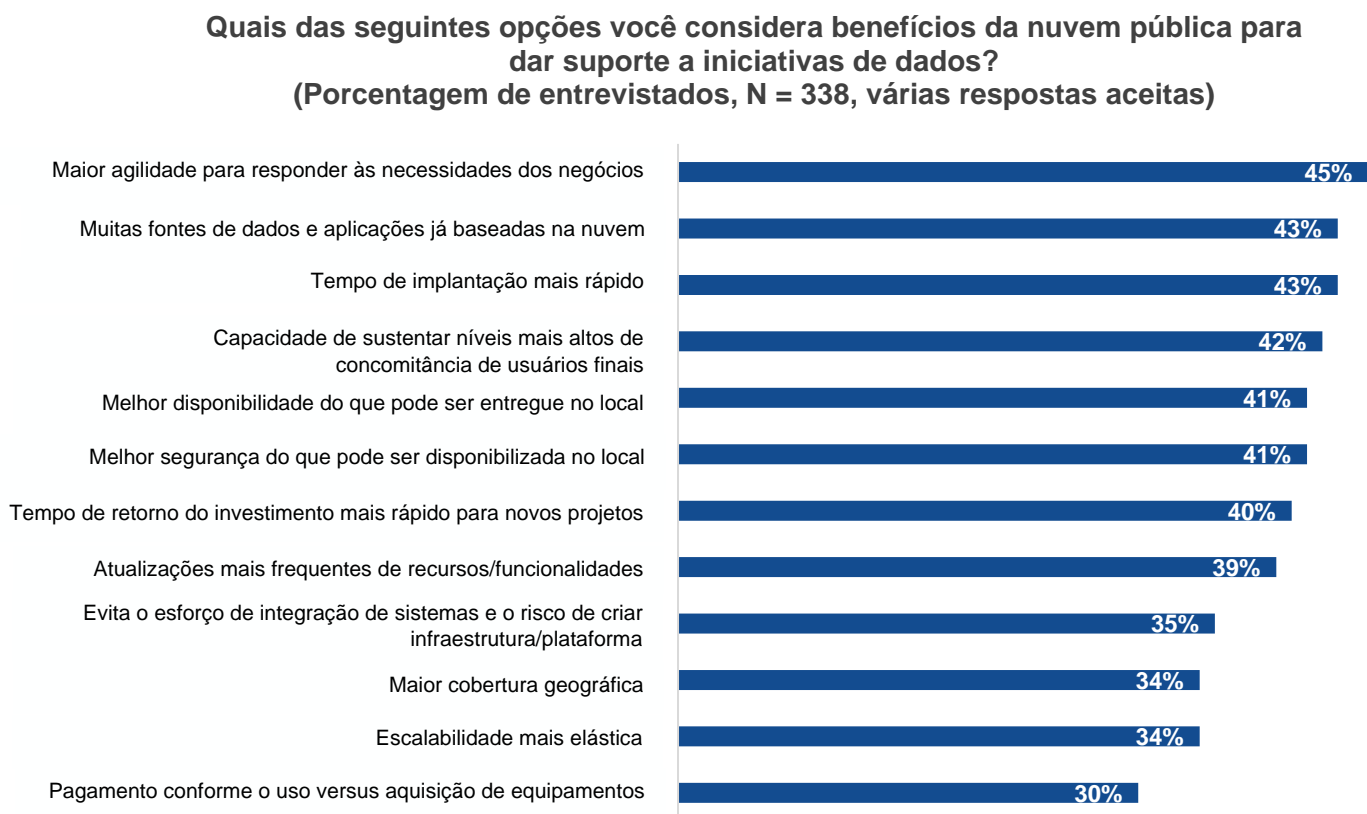
Introdução

Esta validação econômica do Enterprise Strategy Group (ESG) da TechTarget concentra-se nos benefícios quantitativos e qualitativos que as organizações podem esperar obter com o Google Cloud BigQuery em comparação com serviços alternativos de data warehouse corporativos (EDW) baseados em nuvem.

Desafios

Organizações modernas orientadas por dados coletam dados de uma variedade cada vez maior de fontes, dispositivos e locais. Como a linha que antes era clara entre os dados designados para plataformas de dados estruturados, não estruturados, gerenciados e de código-fonte aberto ficou confusa, as organizações estão adotando uma abordagem mais eficiente e consolidando suas estratégias de dados na nuvem. A criação com base em uma nuvem de dados proporciona agilidade, escalabilidade, consolidação simples de fontes de dados e acesso a uma variedade inigualável de ferramentas e serviços de IA e aprendizado de máquina (ML) projetados para extrair percepções e inteligência. Um estudo de pesquisa do ESG descobriu que usar a nuvem pública para dar suporte a iniciativas de dados resulta em maior agilidade para atender às necessidades dos negócios, proximidade a fontes de dados e aplicações já existentes na nuvem, tempos de implantação mais rápidos, capacidade de sustentar níveis mais altos de concomitância de usuário final e maior disponibilidade e segurança (consulte a Figura 1).¹

Figura 1. Principais benefícios de aproveitar a nuvem pública para dar suporte às iniciativas de dados



Fonte: Enterprise Strategy Group, uma divisão da TechTarget, Inc.

¹ Fonte: Relatório do Enterprise Strategy Group, [Cloud Analytics Trends](#), março de 2022. Todas as referências e gráficos de pesquisa do ESG nesta validação econômica foram retirados deste relatório de pesquisa, salvo indicação em contrário.

Nossa pesquisa também mostra que, embora 86% das organizações acreditem que são eficazes em agir com base em novas percepções de dados, apenas 4% dessas organizações extraem informações de seus dados em tempo real, e um impressionante número de 45% das organizações dizem que são necessárias semanas ou mais para obter essa percepção. O data warehouse tradicional no local não é mais ágil ou escalável o suficiente para acompanhar as demandas em constante mudança das modernas organizações orientadas por dados. Uma estratégia que incorpora data warehouses em nuvem, data lakes, IA, ML e transmissão e processamento de dados em tempo real melhorará a qualidade e a pontualidade da percepção. Embora existam muitas soluções de EDW baseadas em nuvem para escolher, a implementação dessas soluções pode ser bem diferente. A escolha da solução de EDW correta requer o planejamento de uma estratégia de nuvem de dados eficaz e uma compreensão não apenas da tecnologia fundamental, mas também dos recursos operacionais e modelos de preços. Com os mandatos da empresa para fazer uso mais eficiente dos recursos, melhorar a visibilidade e otimizar o valor fornecido para cada dólar de gasto em nuvem, é fundamental que as organizações escolham o EDW baseado em nuvem certo.

A solução: plataforma de dados corporativos sem servidor Google Cloud BigQuery

BigQuery é uma plataforma de dados corporativos sem servidor, totalmente gerenciada e baseada em nuvem que oferece suporte a análises de dados em escala de gigabytes a petabytes (consulte a Figura 2). Ela oferece análise de alta velocidade de grandes conjuntos de dados e, ao mesmo tempo, reduz ou elimina investimentos em infraestrutura local ou administradores de banco de dados. O BigQuery aumenta ou diminui o uso de hardware a fim de maximizar o desempenho de cada consulta, adicionando e removendo recursos de computação e storage, conforme necessário.

O BigQuery foi projetado para simplificar a análise e o storage de big data e, ao mesmo tempo, oferecer opções flexíveis e otimizar dados e recursos de computação para economizar custos. Algumas das vantagens específicas do BigQuery para empresas que trabalham com big data incluem:

- **Otimização de custos.** Preveja custos com opções transparentes e flexíveis de cálculo e preço de storage e contenha custos com cotas de projeto e recursos do usuário. As organizações podem optar por pagar taxas de storage lógico (compactado) ou físico (descompactado) com base em suas características de dados e se beneficiar da classificação automática por níveis de storage ativo e de longo prazo para reduzir automaticamente os custos. Para computação, os usuários podem optar por pagar preços flexíveis sob demanda (por terabyte processado) ou edições do BigQuery com preços previsíveis para pagar apenas pelo tempo de computação consumido pelas consultas, permitindo que eles aproveitem os descontos de escalonamento automático e compromisso.
- **Momento de valorização.** Coloque o ambiente de data warehouse on-line de forma rápida, fácil e sem habilidades de administração de banco de dados e sistema de nível especializado, eliminando a infraestrutura e reduzindo o gerenciamento; isso é conhecido como "sem operações" ou "zero operações". A avaliação e os serviços de migração do BigQuery, a ingestão gratuita de dados, a tradução automatizada de SQL e as tecnologias de mecanismo de transmissão minimizam o tempo, o custo e o esforço para obter as primeiras informações provenientes dos dados.
- **Simplicidade.** Conclua todas as principais tarefas relacionadas à análise de data warehouse por meio de uma interface intuitiva sem o incômodo de gerenciar a infraestrutura ou os clusters ou ter que executar funções demoradas de limpeza e otimização. Além disso, ferramentas gratuitas, como o BigQuery Migration Service, ajudam em cada fase da migração, incluindo avaliação e planejamento para execução e verificação, e o Looker Studio oferece visualização gratuita do uso e das tendências dos dados. O Google Cloud gratuito também oferece ferramentas, como Dataform, Dataplex, Data Catalog, Connected Sheets e BI Engine para simplificar o gerenciamento de consultas, canais de dados, metadados e armazenamento em cache para usuários e a organização.
- **Escalabilidade.** Dimensiona de forma transparente a capacidade de computação para atender às necessidades de operação de qualquer porte, sem necessidade de planejamento. Comece com um pequeno volume de dados e amplie até exabytes, dependendo do tamanho, do desempenho e dos requisitos de custo. As reservas de escalonamento automático do BigQuery permitem que as organizações otimizem o custo e o desempenho definindo limites de escalonamento automático para dimensionar automaticamente os slots de acordo com a demanda da carga de trabalho quando necessário e evitar o pagamento excessivo quando os slots não forem necessários.

- **Agilidade.** Aja mais rápido em novas oportunidades de negócios, explore conjuntos de dados para obter novas informações, aproveite os recursos de aprendizado de máquina integrados do BigQuery ML (BQML) usando consultas SQL padrão e inove além das expectativas tradicionais de um data warehouse. O BigQuery Omni oferece uma estratégia ágil de análise de múltiplas nuvens sem precisar mover ou copiar dados entre as nuvens do Google, AWS e Azure. Analise dados de maneira nativa com base em diversos storages, como Cloud SQL, Cloud Storage, Amazon S3, Azure Blob Storage, Google Drive, e um lakehouse com suporte a planilhas Google.
- **Velocidade.** Inclua, consulte e exporte conjuntos de dados em tamanho de gigabyte para petabyte com velocidades impressionantes usando o Google Cloud Platform como infraestrutura de nuvem subjacente.
- **Confiabilidade.** Garanta disponibilidade ininterrupta e tempo de atividade constante em execução no Google Cloud Platform com replicação geográfica em data centers do Google e um contrato de nível de serviço (SLA) líder de 99,99% em tempo de atividade, em comparação com 99,9% para a maioria dos outros provedores.
- **Segurança.** Proteja e controle o acesso a projetos e conjuntos de dados criptografados por meio do gerenciamento de identidade e acesso (IAM) em toda a nuvem do Google e da opção de usar chaves de criptografia gerenciadas pelo cliente (CMEK) para manter o controle sobre os dados criptografados.

Figura 2. Google Cloud BigQuery



Fonte: Enterprise Strategy Group, uma divisão da TechTarget, Inc.

O BigQuery realiza o escalonamento automático; ele identifica os requisitos de recursos para cada consulta a fim de concluir com rapidez e eficiência e fornece esses recursos para atender à demanda. Recursos e capacidade ociosos podem ser usados por outros departamentos ou divisões dentro de uma organização. Assim que a carga de trabalho for concluída, o BigQuery liberará ou realocará automaticamente esses recursos para serem disponibilizados a outros projetos e a outros usuários. Ao mesmo tempo em que transfere dados e processa esses dados para obter resultados, o BigQuery oferece velocidades incríveis, mesmo em escala de petabytes.

Para maior durabilidade dos dados, o BigQuery oferece alta disponibilidade e confiabilidade por meio de replicação geográfica totalmente transparente para seus usuários, sem a necessidade de obter os recursos físicos e espaço para abrigar tudo. O BigQuery oferece backup automático e snapshots de conjuntos de dados, permitindo a restauração em determinado momento no tempo e garantindo que os controles de serviço da nuvem privada virtual (VPC) e a criptografia de dados com a opção de CMEK (em trânsito e em repouso) sejam integrados para fornecer níveis aprimorados de segurança.

Por fim, o BigQuery permite que as organizações enfrentem os desafios de custo e complexidade associados à criação e manutenção de uma infraestrutura de big data rápida, escalável e resiliente. Ao aproveitar a abordagem baseada em nuvem do BigQuery, o tempo e o custo tradicionalmente dedicados à proteção de dados e à garantia do tempo de atividade são praticamente eliminados. Com o Google Cloud lidando com escalabilidade, replicação, proteção e recuperação, as organizações podem se concentrar mais em obter informações valiosas, em vez do gerenciamento de infraestrutura.

O BigQuery oferece uma plataforma unificada para processamento de dados, transmissão, governança e recursos de IA/ML abrangentes. Com suporte a vários formatos de dados (incluindo estruturados e não estruturados), software de código aberto e serviços de parceiros, ele permite a execução de um ciclo de vida completo dos dados. Os clientes do BigQuery aproveitam sua flexibilidade utilizando várias linguagens de codificação, como SQL, Python, Spark e linguagem natural, além de análises entre nuvens. Além disso, os recursos de IA generativa (GenAI) integrados do BigQuery agilizam o tempo de obtenção de informações oferecendo criação e conclusão de código, bem como um bate-papo no console. O ML do BigQuery permite que as organizações desbloqueiem casos de uso de GenAI para dados estruturados e não estruturados usando SQL simples.

Comparação de BigQuery com serviços alternativos de EDW baseados em nuvem

O ESG comparou a solução de EDW sem servidor do BigQuery com alguns serviços alternativos de EDW oferecidos pela AWS, Snowflake e Azure/Databricks. Embora todas essas ofertas ofereçam economia de custos significativa, redução da complexidade e mais agilidade nos negócios em comparação com uma solução de EDW no local, há algumas diferenças significativas entre as ofertas (consulte a Figura 3).

Amazon Redshift. Redshift é o serviço de data warehouse totalmente gerenciado e baseado em nuvem da AWS. Como um cluster local, o Redshift é baseado no conceito de nós (nós virtuais) que devem ser implementados, configurados e gerenciados. O gerenciamento é muito simplificado, pois não há hardware para administrar e manter fisicamente. Para dimensionar a implantação, nós semelhantes de uma capacidade fixa de computação e storage são adicionados simultaneamente, às vezes resultando no provisionamento de mais recursos de computação ou storage para atender aos requisitos um do outro. Os clusters do Redshift exigem planejamento e dimensionamento, configuração e acesso por meio de um "nó líder" e atualizações programadas várias vezes por ano. Embora a AWS ofereça muitas opções de preços flexíveis, para obter maior economia, as organizações devem pagar três anos de serviço antecipadamente, vinculando-as à solução do fornecedor e ao tipo de instância específico que escolheram. A AWS lançou recentemente uma versão sem servidor do Redshift, mas no momento da publicação, não conseguimos coletar informações suficientes relacionadas ao dimensionamento e à administração para incluir com credibilidade neste relatório.

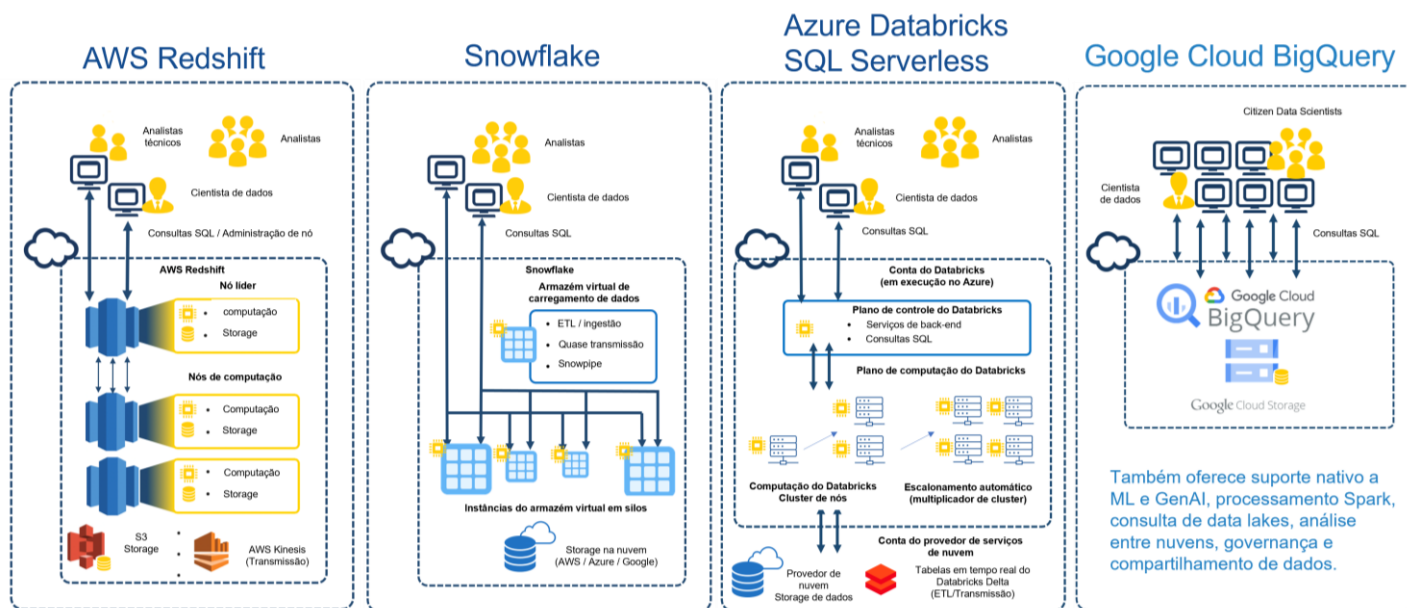
Snowflake. O Snowflake é um data warehouse gerenciado como um serviço (DWaaS) que pode ser implantado na infraestrutura do Google Cloud, AWS ou Azure. O Snowflake não requer gerenciamento de hardware físico ou virtual, instalação de software ou manutenção. O Snowflake também separa os recursos de computação e storage e usa uma arquitetura de processamento paralelo em grande escala (MPP) nos bastidores. No momento da implantação, os clientes devem selecionar entre data warehouses virtuais pré-configurados em vários tamanhos (pequeno, médio, grande, extragrande etc.), que são precificados com base nas "horas de crédito" consumidas. O custo por crédito varia com base na região, no serviço de nuvem e no nível de suporte. Normalmente, os armazéns virtuais são isolados com base em unidades de negócios e podem ser pausados e retomados automaticamente para limitar o custo e serem redimensionados quando necessário. O Snowflake também sugere o uso de um armazém virtual dedicado para lidar com o carregamento e a ingestão de dados para evitar a contenção de recursos necessários para consultas e oferece serviços alternativos com vantagens e desvantagens de custo e confiabilidade adicionais.

Computação do Azure Databricks SQL Serverless. Esta é uma versão de computação sem servidor gerenciada da arquitetura da plataforma Databricks criada no Apache Spark que é executada em clusters de computação em uma conta do Databricks na infraestrutura fornecida pelo Azure. O plano de controle do Databricks fornece um aplicativo da Web para gerenciar consultas SQL, histórico e endpoints, que enviam consultas a armazéns SQL sem servidor que são executados no plano de computação sem servidor em uma rede privada usando instâncias de computação fornecidas na conta Databricks. Os clientes também pagam pelo storage na nuvem necessário

para armazenar o data lakehouse e os dados consumidos pelo Databricks Filesystem (DBFS) ou para armazenar resultados de consulta. O lakehouse do Databricks é baseado no formato de tabela do Delta Lake, que permite o uso do Databricks para processamento de dados em tempo real e geração de relatórios de inteligência de negócios (BI).

EDW sem servidor do BigQuery. A solução BigQuery do Google Cloud é totalmente sem servidor, autoescalável, com automanutenção e autoajuste. Não há nós ou armazéns virtuais para planejar, configurar ou escalar. A complexidade de dimensionar, gerenciar e manter a infraestrutura física é tratada nos bastidores pelo Google, eliminando a carga da organização. Os usuários finais obtêm o benefício de todos os recursos otimizados e ajustados automaticamente que funcionam simultaneamente. Os clientes podem pagar pela quantidade total de dados processados por mês ou optar por economizar com edições previsíveis do BigQuery com pagamento conforme o uso, com base no número de "slots" necessários (efetivamente, a quantidade reservada de recursos paralelos que são disponibilizados para a execução de consultas). Embora os compromissos de slots possam ser alocados a um departamento específico para processar consultas, slots não utilizados são automaticamente alocados a outros departamentos redistribuindo a capacidade ociosa e são disponibilizados para escalonamento automático para lidar com picos, ajudando a otimizar a utilização de slots. O BigQuery também é compatível com aprendizado de máquina baseado em SQL nativo, análise de sistemas de informações geográficas (GIS), tabelas BigLake para consultar data lakes e o BigQuery Omni para análise em nuvem cruzada. Dessa maneira, todos os usuários se beneficiam do poder da análise avançada sem a necessidade de integração com outras soluções ou de migração de dados entre tecnologias e nuvens.

Figura 3. Comparação funcional de soluções de EDW baseadas em nuvem



Fonte: Enterprise Strategy Group, uma divisão da TechTarget, Inc.

Validação econômica do Enterprise Strategy Group

O processo de validação econômica do Enterprise Strategy Group (ESG) é um método comprovado para entender, validar, quantificar e modelar as propostas de valor econômico de um produto ou solução. O processo aproveita as principais competências do ESG em análise de mercado e do setor, pesquisa voltada para o futuro e validação técnica e econômica. Para validar as suposições e os custos incluídos na análise, o ESG realizou entrevistas detalhadas e analisou estudos de caso de organizações que migraram anteriormente suas operações de uma solução de EDW local existente para o BigQuery. Isso nos ajudou a entender melhor e quantificar como a mudança para o BigQuery afetou suas organizações. Usamos nossas descobertas para criar um modelo econômico detalhado que compara os custos e os benefícios esperados do BigQuery com as soluções baseadas em nuvem da concorrência.

Visão geral do valor econômico: comparação do BigQuery com soluções alternativas de EDW baseadas em nuvem

A análise econômica do ESG revelou que o BigQuery pode oferecer economias operacionais e de capital significativas, bem como benefícios tangíveis em comparação com outras soluções de EDW baseadas em nuvem. O ESG descobriu que o BigQuery proporcionou aos clientes economias e benefícios significativos nessas categorias:

- **Eliminação de investimento inicial e planejamento.** O projeto sem servidor do BigQuery é cobrado mensalmente com opções de preços flexíveis e previsíveis de computação e storage que podem se adequar melhor à necessidade de diferentes organizações e cargas de trabalho. Essas opções eliminam a necessidade de pagar meses ou até mesmo anos com antecedência para reduzir os gastos com nuvem e eliminar adivinhações e o planejamento necessários para dimensionar os recursos corretamente. Isso proporciona à organização mais agilidade nos negócios quando se trata de custos de serviços em nuvem.
- **Redução das despesas operacionais.** O BigQuery elimina a necessidade de gerenciar nós virtuais do EDW, bem como a necessidade de monitorar, solucionar problemas, atualizar, ajustar e planejar o crescimento. O BigQuery Storage é auto-otimizado em termos de custo, não são necessários patches e manutenção, e a equipe de suporte é bem treinada e ágil. Isso permite aos administradores mais tempo para se concentrarem em outras áreas da empresa.
- **Mais agilidade nos negócios e redução do custo da administração diária.** O BigQuery pode ser ampliado ou diminuído, conforme necessário, para atender às demandas de negócios em constante mudança, permitindo que as organizações atuem rapidamente em novas oportunidades sem a necessidade de planejar requisitos de configuração, pausar bancos de dados ou acelerar instâncias de armazém dedicadas para cada organização. A solução também ajuda a eliminar ou reduzir o tempo gasto na administração do banco de dados; gerenciamento de extração, transformação e carregamento (ETL); e nova modificação de esquema. O BigQuery oferece recursos nativos de IA/ML e fácil integração com GenAI. Ele também suporta consultas nativas de data lakes com BigLake e uma estratégia de análise de múltiplas nuvens com o BigQuery Omni.
- **Menor custo e complexidade da execução de cargas de trabalho de ML e casos de uso de GenAI.** O mecanismo de inferência ML do BigQuery permite o uso de SQL simples no BigQuery para executar previsões usando modelos hospedados remotamente e modelos básicos do Google para dados estruturados no storage gerenciado e dados não estruturados do BigQuery no Google Cloud Storage. Isso elimina a necessidade de movimentação de dados, reduzindo a complexidade, os custos e os riscos de segurança para os clientes, ao mesmo tempo em que reduz a necessidade de depender de recursos qualificados com conhecimento de linguagens de programação avançadas.

Eliminação de investimento inicial e planejamento conforme o porte da implantação

A migração de um EDW local para qualquer solução EDW baseada em nuvem elimina a necessidade de fazer um grande investimento inicial em infraestrutura física e ajuda a evitar o provisionamento excessivo de hardware físico de computação e storage para crescimento e picos de demanda. No entanto, nem todas as soluções de EDW baseadas em nuvem evitam completamente a necessidade de fazer investimentos iniciais, a necessidade de planejar o porte da implantação ou a necessidade de superprovisionamento para lidar com picos de carga de trabalho (consulte a Tabela 1).

- Eliminação do investimento inicial.** O BigQuery não requer investimento inicial para atingir o maior nível de economia para recursos de computação. Os descontos oferecidos por meio de compromissos de slots das edições do BigQuery fazem parte de um contrato, com o custo pago mensalmente. O Snowflake oferece descontos para pagamentos antecipados e custos de storage mais baixos para aqueles que reservam capacidade, mas não requer um investimento inicial. (No entanto, o cliente é obrigado a pagar pelo storage não utilizado.) O AWS Redshift exige contratos anuais e investimento inicial para atingir o maior nível de economia. Essa quantia deve ser paga com um ou três anos de antecedência, mesmo que o valor do serviço não seja percebido até meses ou anos depois. Da mesma forma, o Azure Databricks SQL Serverless oferece descontos por pré-compra de Databricks Commit Units (DBCUs) como uma única compra com um ou três anos de antecedência. Como qualquer investimento de capital, as organizações devem ter certeza de levar em consideração o custo adicional de capital ao longo do tempo ao comparar os investimentos iniciais com os custos mensais de compromisso sob demanda.
- Sem planejamento do tamanho da implantação da infraestrutura.** Com o BigQuery, os clientes não precisam planejar e dimensionar a configuração para atender às suas necessidades. O BigQuery é completamente sem servidor e os clientes não precisam planejar nem ajustar a infraestrutura de suporte. Com o AWS Redshift, o tamanho e a quantidade das instâncias necessárias para lidar com a ingestão e a carga de trabalho devem ser previstos e dimensionados como se fossem nós locais. Com o Azure Databricks SQL Serverless e o Snowflake, os usuários devem determinar a combinação de storage e o tamanho do cluster de nós de computação ou do data warehouse virtual que melhor se adapta às necessidades de cada unidade de negócios. O Snowflake também recomenda um armazém dedicado usado para carregar dados.
- Não há necessidade de aumentar a implantação.** O BigQuery sob demanda não requer monitoramento da utilização de capacidade ou crescimento do ambiente quando os limites são atingidos, e os compromissos de slots das edições podem ser facilmente expandidos sempre que necessário. As edições do BQ aproveitam o escalonamento automático para ampliar até um número máximo de slots predefinidos durante os tempos de pico e, em seguida, retornar aos níveis normais quando não forem necessários. As implantações do AWS Redshift podem ser ampliadas, mas isso requer não apenas sobrecarga de gerenciamento, mas também algum tempo de inatividade temporário do cluster à medida que as instâncias são adicionadas. O dimensionamento de recursos para o Snowflake e o Azure Databricks SQL Serverless é relativamente fácil, mas requer algum grau de planejamento e atividade administrativa. Essas ofertas também oferecem dimensionamento automático de armazéns, mas essas ofertas devem ser dimensionadas com a granularidade do tamanho do cluster original (múltiplos) e exigir um tempo maior para implantar, aquecer e limpar após o uso, custando mais ao usuário. A Figura 4 ilustra a vantagem do escalonamento automático baseado em slot para atender melhor às necessidades de cargas de trabalho em constante mudança, de uma perspectiva de agilidade e economia de custos. O Google Cloud estima que, com o escalonamento automático granular, algumas organizações podem reduzir seus gastos comprometidos com computação em 30% a 40%.

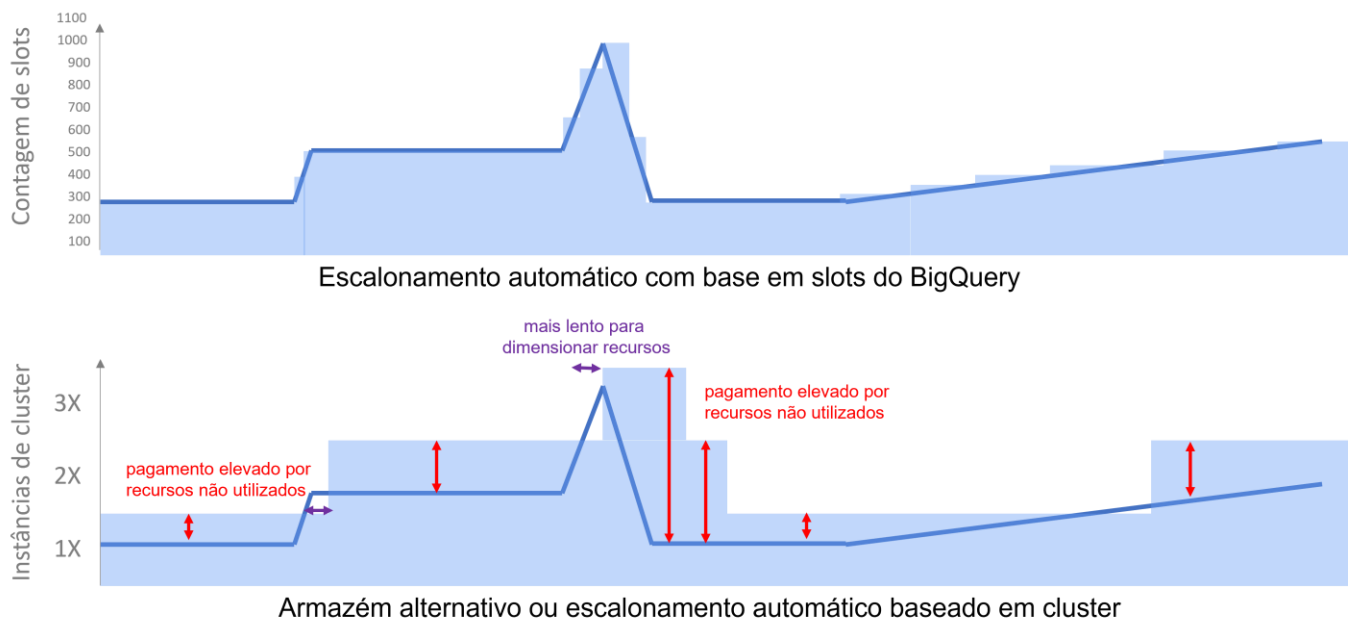
"Com o Google, não estamos presos a nenhum prazo ou obrigados a pagar antecipadamente para economizar, como outros contratos de nuvem, o que é ótimo para nós à medida que ampliamos nossos recursos e nossas necessidades mudam. Posso alterar as reservas e os modelos de preços todos os meses, conforme necessário. Esse modelo nos dá muito mais flexibilidade do que qualquer outra coisa que eu tenha usado."

- Não há necessidade de superprovisionamento.** A infraestrutura do BigQuery utiliza todos os recursos disponíveis para processar consultas e não precisa ser superprovisionada para lidar com picos inesperados na carga de trabalho. Os clusters de computação AWS Redshift, Snowflake Virtual Warehouses e Azure Databricks SQL Serverless definem um conjunto fixo de recursos que estão disponíveis para as consultas serem executadas, além de gerenciar processos de ingestão, recluster e limpeza. Nesse sistema, uma

"Assim que implementamos os preços de reservas do BigQuery, começamos a ver uma mudança drástica no faturamento. Sem fazer quaisquer outras mudanças, reduzimos o nosso gasto mensal em 30%."

implantação de configuração muito pequena afeta negativamente as consultas, levando as organizações a dimensionar e pagar por implantações de linha de base para lidar com o pior cenário para cargas de trabalho de pico. O BigQuery permite que qualquer recurso em qualquer departamento da organização use os slots fixos disponíveis e a capacidade compartilhada, enquanto os departamentos de outras soluções não podem se beneficiar de computação ou storage ocioso ou com superprovisionamento em clusters de outros departamentos, limitando ainda mais a eficiência dos recursos com base no número de clusters implantados. Embora as implantações em vários clusters no Azure Databricks SQL Serverless e Snowflake possam ajudar a superar as limitações de dimensionamento, elas devem aumentar no mínimo 2x os recursos, contribuindo ainda mais para a ineficiência do superprovisionamento em muitos casos e estão limitadas a um número máximo de clusters.

Figura 4. Escalonamento automático baseado em slot versus baseado em cluster



Fonte: Enterprise Strategy Group, uma divisão da TechTarget, Inc.

- Custo reduzido de migração.** Embora cada fornecedor ofereça ferramentas e serviços para ajudar a migrar dados para seu banco de dados na nuvem, os clientes relataram que a migração para o BigQuery era muito rápida e simples, com ferramentas gratuitas fornecidas pelo Google Cloud, enquanto outras ferramentas e serviços complementares de fornecedores apresentavam um custo adicional. O Google Cloud fez investimentos significativos em aquisições de tecnologia, melhorias e ferramentas para migração e integração. A ferramenta de avaliação automatizada do Google ajuda a entender os ativos e priorizar tarefas. A ingestão de dados em lote é gratuita. A tradução SQL interativa e em lote por meio da aquisição do CompilerWorks reduz o tempo e o esforço manual e reduz o risco de erros. A ferramenta de validação de dados do Google Cloud valida tabelas de origem e destino. Ao reduzir os recursos necessários para planejar, traduzir, ingerir, validar e treinar clientes em um novo esquema, o BigQuery pode limitar o custo operacional inicial e acelerar o tempo de retorno ao valor comercial.

Tabela 1. Comparação de soluções de EDW baseadas em nuvem: Investimento Inicial, planejamento e agilidade

	Google BigQuery	AWS Redshift	Snowflake	Azure Databricks SQL Serverless
Investimento inicial (Pago antecipadamente pelos serviços prestados)	Nenhum. Contratos de tarifa fixa ainda são pagos mensalmente, não antecipadamente.	Contratos de um ou três anos com investimentos iniciais necessários para receber taxas competitivas.	Nenhum para computação. Capacidade de storage reservada necessária para o menor preço, seja o storage usado ou não.	Pago conforme o uso com uma opção de descontos por uso comprometido (storage). É necessário comprar o crédito antecipado (DBCU) para obter a melhor economia.
Dimensionamento e planejamento	Nenhum planejamento necessário. O BigQuery lida com todas as alocações de slots. Não há necessidade de gerenciar a computação.	É necessário prever os tamanhos corretos de instâncias e as instâncias reservadas limitam a agilidade.	Preveja o tamanho do armazém virtual necessário para cada unidade de negócios.	É preciso prever o tamanho do cluster sem servidor e ligar manualmente. É preciso encerrar clusters ociosos ou incorrer em custos (pode ser encerrado automaticamente).
Agilidade/ crescimento	Dimensionamento automático de impacto zero. Nenhum ajuste é necessário para preços sob demanda. É possível definir limites simultaneamente (padrão = 100). Slots adicionais podem ser aumentados sempre que necessário.	O redimensionamento do tamanho do cluster requer tempo de inatividade do cluster. Limitado a 15 consultas simultâneas por cluster. É possível dimensionar automaticamente para 10 clusters, mas limitado a 50 consultas em fila.	Dimensione adicionando novos armazéns virtuais ou aumentando o tamanho de armazéns virtuais individuais. Armazéns com vários clusters devem aumentar até 2x. Limite de 8 consultas simultâneas por armazém, dimensionamento automático para 10 armazéns (80 consultas simultâneas).	Dimensionamento automática adicionando vários clusters de nós de tamanho selecionado. Limite de 10 consultas simultâneas por cluster. Leva tempo para dimensionar e aquecer novas cópias de clusters.
Superprovisionamento para crescimento de capacidade ou picos de desempenho	Nenhum.	Deve realizar o superprovisionamento de instâncias suficientes para lidar com o pior cenário ou aumentar a sobrecarga operacional e o tempo de inatividade de nós de aumento e diminuição, conforme necessário.	Deve realizar o superprovisionamento de instâncias suficientes para lidar com o pior cenário ou aumentar a sobrecarga operacional e o aumento e a diminuição de armazéns, conforme necessário.	Suporta dimensionamento automático, limitando a necessidade de superprovisionamento, mas é necessária a previsão de nós mínimos e máximos de trabalho para cada trabalho.
Custos de storage	Escolha de storage físico (compactado) ou lógico para otimizar o custo.	Storage e computação separados estão disponíveis apenas em RA3 nós com storage gerenciado. Os tipos de nós mais antigos devem ser configurados com capacidade de storage de SSD ou HDD interno.	Storage microparticionado com capacidade de dados de limite de compactação que é verificada. O serviço de otimização de pesquisa pode pré-indexar algumas consultas por um custo adicional.	Custo de storage pago à Azure. O Parquet e o Delta Lake comprimem o storage para reduzir os custos.

Fonte: Enterprise Strategy Group, uma divisão da TechTarget, Inc.

Redução nas despesas operacionais

Eliminar a necessidade de gerenciar e manter o hardware no local gera economias operacionais consideráveis para as organizações que migram para serviços baseados em nuvem. A natureza completamente sem servidor do BigQuery traz vantagens operacionais adicionais, mesmo quando comparado a outros serviços de EDW baseados em nuvem destinados a reduzir esses custos (consulte a Tabela 2).

- **Eliminação do gerenciamento de nós.** O BigQuery é completamente sem servidor e sem estado. O AWS Redshift requer gerenciamento de nó virtual para tarefas como configurar o acesso à rede e VPNs, alocar recursos, atualizar software, manter o acesso aos recursos e aumentar e reduzir o cluster. O Snowflake e o Azure Databricks SQL Serverless são mais simples do que o Redshift, mas ainda exigem dimensionamento e gerenciamento manuais para vários armazéns virtuais ou clusters de computação. Um cliente afirmou: "Tínhamos aproximadamente 60 pessoas planejando, gerenciando e mantendo nós de infraestrutura física ou virtual no local e na AWS. Essa foi uma parte significativa da nossa equipe que conseguimos remanejar para outras atividades."
- **Gerenciamento de storage reduzido.** O BigQuery é a única solução de EDW baseada em nuvem entre as quatro opções que oferece storage totalmente gerenciado e com otimização automática. Os dados com mais de 90 dias são movidos automaticamente para um storage econômico visando reduzir os custos em 50%. O BigQuery mantém cópias replicadas de dados para disponibilidade de 11 noves, criptografa todos os dados antes de gravar em disco, dá suporte a clones e snapshots eficientes baseados em delta e permite o reparticionamento de tabelas.
- **Manutenção de cálculo zero.** O BigQuery, o Snowflake e o Azure Databricks SQL Serverless são mantidos sem interrupções para o usuário em segundo plano. Nós do Snowflake mantêm o estado, e o Azure Databricks SQL Serverless e o Redshift têm caches de SSD para dados, portanto, deixá-los inativos pode afetar o desempenho. O AWS Redshift requer que as janelas de manutenção manual sejam programadas, e as atualizações podem resultar em tempo de inatividade.
- **Estratégia de nuvem de dados unificada.** As organizações modernas se beneficiam do uso da estratégia de nuvem de dados unificada do Google Cloud para aproveitar ao máximo seus dados e reduzir ainda mais os custos e a complexidade, integrando e otimizando o BigQuery com outros produtos de dados do Google Cloud. O Google Cloud BigLake permite que as organizações unam data lakes e armazéns em um único conjunto de dados. Isso simplifica a governança de dados e o gerenciamento do ciclo de vida e reduz cópias duplicadas de dados, ao mesmo tempo em que simplifica e acelera a ingestão e a transformação de dados. O BigQuery Omni permite que os usuários acessem dados em outras nuvens públicas, e o BigSearch fornece recursos de consulta para análise de registros baseada em JSON.

"Nossa principal dor de cabeça com nosso data warehouse em nuvem anterior era manter o banco de dados e ajustar e configurar o cluster. Com o BigQuery, podemos nos concentrar em nossos canais, nos dados e em nossos resultados."

Tabela 2. Comparação de soluções de EDW baseadas em nuvem: Despesas operacionais

	Google BigQuery	AWS Redshift	Snowflake	Azure Databricks SQL Serverless
Gerenciamento de nó virtual	Nenhum; completamente sem servidor.	Nó de gerenciamento necessário com configuração manual e dimensionamento de nós.	Sem servidor, mas os usuários podem precisar gerenciar data warehouses em silos, e um depósito dedicado para carregar dados é sugerido.	Nenhum cluster a ser gerenciado, mas os usuários devem dimensionar e iniciar clusters, gerenciar o acesso dos usuários, converter dados em formato delta (o que requer computação extra) e gerenciar políticas de dimensionamento/finalização.
Gerenciamento de storage	O storage otimizado automaticamente move os dados para um storage econômico.	Migração manual e envelhecimento do storage.	Deve gerenciar a movimentação de dados e decidir entre storage caro e rápido e storage econômico de capacidade.	É necessário gerenciar as opções de storage (storage Azure Blob ou Data Lake). O formato de storage requer otimização periódica do storage.
Manutenção de EDW / nó	Gerenciado pelo Google Cloud em segundo plano, sem tempo de inatividade.	Atualizações manuais de nós durante o tempo de inatividade programado.	Gerenciado pelo Snowflake em segundo plano, sem tempo de inatividade.	Gerenciado pelo Databricks em segundo plano, sem tempo de inatividade.

Fonte: Enterprise Strategy Group, uma divisão da TechTarget, Inc.

Maior agilidade nos negócios e redução no custo da administração diária do EDW

A administração diária do EDW geralmente consiste em tarefas como gerenciar o processo ETL entre sistemas; configurar, gerenciar e monitorar a plataforma e o banco de dados; solucionar problemas; controlar o acesso; manter a segurança; desenvolver novos serviços; colaborar com analistas de negócios; fornecer dados para relatórios e painéis; e integrar aplicativos e outros serviços de nuvem. O ESG descobriu que, quando comparado com uma solução no local, o BigQuery e outras soluções baseadas em nuvem podem eliminar algumas tarefas completamente e tornar outras tarefas mais simples e rápidas por meio da automação e da integração. O BigQuery oferece mais simplicidade de gerenciamento, mesmo quando comparado a outros serviços baseados em nuvem (consulte a Tabela 3).

- **Eliminação da necessidade de monitorar e redimensionar recursos virtuais.** Os administradores do BigQuery não precisam gastar tempo monitorando ou dimensionando os recursos virtuais do sistema (nós, unidades de computação ou armazéns virtuais). Como os limites de simultaneidade do BigQuery são flexíveis, eles podem ser facilmente aumentados quando necessário. Para aqueles que desejam monitorar o ambiente do BigQuery, os painéis de monitoramento integrados permitem uma percepção do desempenho das consultas, tendências, faturamento e auditorias do sistema, com a capacidade de criar alertas automatizados com base em limites.
- **Não há necessidade de executar tarefas periódicas de otimização e limpeza de EDW.** O BigQuery realiza o reagrupamento automático de tabelas para obter o desempenho ideal e elimina a necessidade de realizar manutenção periódica, como processos de limpeza, que podem consumir recursos e afetar negativamente o desempenho. Algumas das ofertas alternativas exigem processos manuais ou automáticos de reclustering e limpeza que são executados em segundo plano, e todos eles consomem recursos de computação que devem ser pagos, e as operações podem afetar o desempenho de consultas ativas.

- **Maior agilidade nos negócios.** Escalabilidade responsiva significa que as consultas podem ser concluídas mais rapidamente, decisões informadas podem ser tomadas mais rapidamente e recursos para satisfazer oportunidades de negócios urgentes estão disponíveis o tempo todo. Os recursos não precisam ser alocados ou expandidos para atender às necessidades de novas oportunidades. Os clientes do BigQuery podem compartilhar compromissos de slots entre projetos para lidar com alterações na demanda de carga de trabalho. Os clientes do BigQuery relataram que eram muito mais ágeis e capazes de agir com base em oportunidades muito mais rapidamente, sem impactos nos negócios.
- **ETL simplificado e gerenciamento de fontes de dados.** Os clientes com quem o ESG conversou informaram que as opções flexíveis permitem que eles organizem dados no local ou no Google Cloud e armazenem dados em uma variedade de opções no Google Cloud – Cloud Spanner, BigTable, Cloud Storage etc. – ou de fontes de dados externas. Essas opções flexíveis permitem maiores possibilidades e recursos por meio do uso de outros produtos do Google Cloud. O Google Cloud Dataform ajuda a simplificar o projeto e o gerenciamento do canal de dados. Algumas opções alternativas existem no ecossistema AWS, mas Snowflake e Azure Databricks SQL Serverless não dependem de integrações com serviços de nuvem pública.
- **Ingestão simplificada de dados.** A ingestão de dados é um dos pontos fortes do BigQuery em comparação com as soluções de EDW alternativas baseadas em nuvem. O ESG analisou os resultados de testes confiáveis de terceiros que mostraram que o BigQuery atingiu uma pontuação de 1,9x a 2,4x a mais do que as soluções alternativas em recursos de ingestão de dados. A ingestão de dados em lote é gratuita e não consome nenhuma capacidade de consulta do cliente no BigQuery. O AWS Redshift e o Snowflake exigem configurações e tarefas muito mais complexas para gerenciar as operações de ingestão, e as operações de ingestão aproveitam os recursos de computação do usuário, basicamente exigindo o provisionamento excessivo dos recursos de computação para atender à demanda da carga de trabalho. O serviço Snowflake sugere a implantação de um armazém virtual dedicado para carregamento e ingestão de dados, o que aumenta o custo e a complexidade do gerenciamento. O serviço Snowpipe da Snowflake pode facilitar a ingestão, mas aumenta o custo e os trabalhos consomem recursos de computação adicionais.
- **Tarefas legadas simplificadas relacionadas ao EDW.** Os clientes do BigQuery informaram que economizavam tempo eliminando backups e tarefas de manutenção do sistema, simplificando processos como integração, gerenciamento de prioridades de carga de trabalho e manutenção de partições.
- **Controle de acesso e segurança simplificados.** O Google Cloud Deployment Manager pode ajudar a automatizar a criação de funções personalizadas de IAM, que fornecem gerenciamento de projetos simplificado e controle de acesso por meio da flexibilidade para gerenciar rapidamente o acesso a recursos por função funcional, organização, estágio do ciclo de vida de dados ou projeto. O BigQuery elimina a necessidade de gerenciar usuários individuais ou organizar e conceder permissões, além de ter recursos de segurança e criptografia integrados com chaves gerenciadas pelo BigQuery ou pelo cliente.

"Descobrimos que o Redshift era complicado de implementar e dimensionar, o BigQuery foi muito mais fácil. Ele nos dá uma visão melhor de nossos projetos e oferece custos mais previsíveis para o que estamos executando."

Tabela 3. Comparação de soluções de EDW baseadas em nuvem: agilidade aprimorada e gerenciamento simplificado do EDW

	Google BigQuery	AWS Redshift	Snowflake	Azure Databricks SQL Serverless
Planejar e dimensionar requisitos de recursos virtuais para novas cargas de trabalho	Não é necessário dimensionamento adicional.	Dimensionar instâncias da AWS, considerando computação, memória e storage local.	Dimensionar um novo armazém virtual e em silos ou dimensionar o armazém existente.	Preveja os tipos de instância de IaaS de linha de base e o tamanho inicial do cluster.
Provisionar cluster/armazéns virtuais/rede	Nenhum provisionamento é necessário.	Provisionar VPC, rede, acesso de usuário e recursos de storage em novas instâncias. Tempo de inatividade necessário.	Provisionamento simples.	Provisionamento simples.
Tarefas periódicas de otimização e limpeza do EDW	Reclustering automático de tabelas em segundo plano. Não é necessário realizar a limpeza de processos. Sem faturamento para computação.	Otimização automática da tabela. É necessária uma operação de limpeza frequente em horários fora de pico para limpar as tabelas e manter um desempenho consistente. Cobrado pela computação consumida.	Reclustering automático e operações de limpeza sem impactos na produção. Cobrado pela computação usada.	O serviço gerenciado executa o processo de compactação e limpeza para otimizar o desempenho do storage e da consulta. Cobrado pela computação usada.
Gerenciar/monitorar/dimensionar recursos virtuais	Não é necessário.	Monitore o desempenho do cluster e realize scale up/redução conforme necessário. Instâncias reservadas não podem ser reduzidas.	Monitore o desempenho do cluster e realize scale up/redução/scale out conforme necessário. A dependência do armazenamento em cache pode resultar em desempenho imprevisível.	Monitore o desempenho do trabalho anterior e ajuste os recursos iterativamente ou use o escalonamento automático.
Gerenciar patches de acesso do usuário/segurança	Controles de IAM simplificados; acesso baseado em funções com aplicação automática de patches.	Gerencie usuários e segurança para cada cluster. Aplicação de patches durante janelas programadas.	Permissões simplificadas baseadas em funções e no nível de objeto com aplicação automática de patches.	Gerencie usuários e segurança para cada cluster. Aplicação automática de patches.
Gerenciar carregamento de dados/ingestão/transmissão de dados	Opções flexíveis de ingestão, transformação e transmissão de dados. O mais alto desempenho de rede. ETL/ELT com Cloud Data Fusion.	O carregamento e a transmissão de dados complexos exigem o serviço Kinesis. A ingestão consome recursos de computação alocados. ETL/ELT com AWS Glue.	Sugere-se que um armazém separado esteja em execução para carregamento de dados. O serviço Snowpipe funciona quase em tempo real.	Gerencie clusters separados e crie tarefas para executar canais e transformações de dados. Transmissão com Delta Lakes e Kafka.
Integre com outras ferramentas e serviços baseados em nuvem	Projetado para integração com muitos serviços do Google Cloud, incluindo DataFlow, DataProc, CloudDB, PubSub e Planilhas, exigindo o mínimo de sobrecarga de gerenciamento. Análise de múltiplas nuvens com o BQ Omni.	Integração manual com serviços da AWS; alguma sobrecarga de gerenciamento e configuração.	A integração com serviços de outros fornecedores requer sobrecarga operacional e de gerenciamento.	Integrações limitadas com algumas ferramentas de nuvem e código aberto para BI, ETL e IDE exigem sobrecarga operacional e de gerenciamento.
Integrar com IA/ML	Recursos integrados para criar e executar modelos de ML. Suporte a casos de uso de GenAI e ML avançado por meio de integrações de IA Vertex.	Recursos de ML via Sagemaker. GenAI com Amazon Q.	Recursos limitados de ML; devem usar outras soluções. Várias ferramentas de GenAI na visualização.	A oferta sem servidor não suporta ML

Fonte: Enterprise Strategy Group, uma divisão da TechTarget, Inc.

Considerações sobre migração

A migração de cargas de trabalho de EDW no local para uma solução ou serviço de EDW baseado em nuvem oferece vantagens e oportunidades significativas para economia de custos, mas isso não pode ser conseguido de forma realista da noite para o dia. As organizações devem analisar quais tarefas estão envolvidas e elaborar um plano sólido para decidir quando mover seu EDW para a nuvem.

- **Estratégia de migração para EDW.** As organizações devem considerar se desejam migrar todos os dados de uma só vez e seguir com as operações existentes ou migrar as operações lentamente ao longo de um período de anos, ao mesmo tempo em que eliminam gradualmente o EDW no local. As organizações devem considerar os custos envolvidos na migração em termos de tempo, transferência física de dados, serviços profissionais, possível tempo de inatividade etc.
- **Custos de desenvolvimento e teste.** As organizações devem considerar os custos envolvidos no redesenvolvimento de aplicativos, na transformação de dados em novos esquemas, na regravação de consultas otimizadas, testes, validação, solução de problemas, criação de aplicativos personalizados, novo treinamento de desenvolvedores etc.
- **Reprojeto do processo.** A mudança apresenta a oportunidade de melhorar muitos processos e sistemas. As organizações devem considerar o custo em termos de tempo e dinheiro para fazer alterações no processo ETL, incluindo recursos de transmissão, integração com outros produtos de nuvem e treinamento de equipes de negócios e recursos de TI sobre os novos processos e ferramentas.
- **Custos de servidor e software.** As organizações devem considerar os custos associados a recursos suplementares, como servidores de preparação no local ou baseados em nuvem ou licenças de software e SaaS, sem esquecer de equilibrar esses novos custos com qualquer economia potencial obtida com a migração para a nuvem.
- **Estratégia de múltiplas nuvens.** As organizações devem considerar a capacidade de aproveitar seu data warehouse baseado em nuvem para acessar dados armazenados em qualquer nuvem pública. O BigQuery Omni permite que os dados sejam analisados usando o BigQuery na AWS ou no Azure sem precisar migrar ou copiar dados para o Google Cloud. Isso proporciona economia de custos e maior flexibilidade e agilidade.

Para ajudar a garantir uma migração bem-sucedida, o Google Cloud oferece às organizações uma estrutura de migração prescrita com recursos, financiamento e uma metodologia comprovada.²

Análise do Enterprise Strategy Group

O ESG aproveitou as informações coletadas por meio de material fornecido pelo fornecedor, conhecimento público e do setor sobre economia e tecnologias, e os resultados das entrevistas com clientes para criar um modelo de três anos para o modelo de TCO/ROI que compara os custos e os benefícios de satisfazer os requisitos de EDW de uma organização modelada com o BigQuery em comparação com outras três soluções de EDW baseadas em nuvem. As entrevistas do ESG com clientes que têm experiência com várias soluções de EDW em nuvem e informações de preços publicamente disponíveis, além de diretrizes de dimensionamento, combinadas com nossa experiência e habilidade em modelagem econômica e validação técnica, ajudaram a formar a base para as suposições usadas em nosso cenário modelado.

Para refletir o tamanho e a escala dos clientes entrevistados, o ESG considerou um cenário de uma grande organização comum (mais de 10.000 funcionários) com requisitos para armazenar e analisar aproximadamente 200 TB de dados gerados por várias unidades de negócios. Desses dados, 110 TB foram presumidos como um conjunto de dados de tamanho fixo de tabelas históricas e atualizadas que não expiraram, e o ESG assumiu que 250 GB de dados foram incluídos diariamente com um período de retenção de 365 dias. Nossos modelos previram que, com base nesses parâmetros, cada uma das soluções exigiria aproximadamente 100 TB de storage após 50% de compactação. Presumimos que a organização executou consistentemente uma linha de base média de consultas em toda a organização, o que resultou na necessidade de cerca de 800 slots do BigQuery 24 horas por dia. Além disso, houve alguns períodos pesados de atividade durante o dia em que a necessidade de slots adicionais aumentou em cerca de 1.000 slots, para uma média de cerca de uma hora por dia. Às vezes, consultas de alta prioridade chegaram a até 2.000 slots para períodos curtos e pouco frequentes. (Presumimos que essa média tenha sido de cerca de 5 minutos por dia.)

² Fonte: Google Cloud, [Migrating Teradata and other data warehouses to BigQuery](#), agosto de 2019.

O ESG assumiu que o preço da implantação do BigQuery era calculado aplicando o preço de computação pago conforme o uso da Edição Enterprise, com uma capacidade confirmada de um ano de 800 slots e a capacidade de escalonamento automático de até 2.000 slots adicionais para lidar com demandas de pico. As outras soluções foram dimensionadas usando estimativas conservadoras de relações entre o número de slots fixos do BigQuery e unidades de computação elástica (ECU) (AWS Redshift) e um tamanho da implantação do cluster Azure Databricks SQL Serverless e Snowflake equivalente (mostrado na Tabela 4). Essas relações foram obtidas de boa-fé analisando informações coletadas em campo com base nos testes comparativos de prova de conceito, substituições de soluções e dados publicados. Presumimos que a implantação do Redshift seria executada continuamente. Também presumimos que as configurações do Snowflake e do Azure Databricks SQL Serverless atenderiam às cargas de trabalho equivalentes a 800 slots de base 24/7 e seriam automaticamente redimensionadas de acordo para atingir os níveis de slots aumentados equivalentes, com uma sobrecarga de tempo adicional para acomodar um mínimo de suspensão automática de 10 minutos e concluir os processos de limpeza. Na prática, as soluções competitivas podem exigir ainda mais recursos e/ou tempo de escalonamento automático para lidar com a sobrecarga adicional de recursos necessária para a ingestão de dados, o recluster e os processos de vácuo, que são gratuitos e não têm impacto negativo no desempenho no BigQuery. Para tornar as soluções o mais equivalentes possível de uma perspectiva de recursos para a Edição Enterprise do BigQuery, usamos preços para a Edição Enterprise do Snowflake e o nível Premium do Azure Databricks SQL Serverless.

As suposições de configuração para cada solução são mostradas na Tabela 4.

Tabela 4. Suposições de configuração usadas no cenário modelado de três anos do ESG

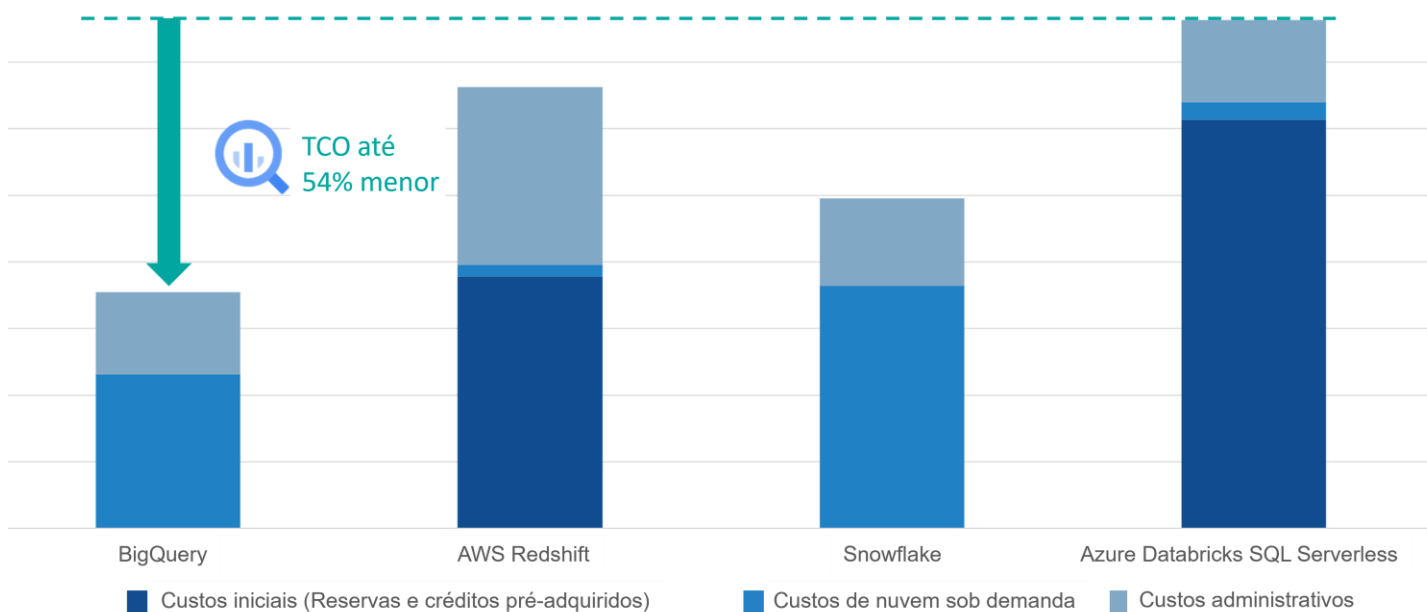
	Google BigQuery	AWS Redshift	Snowflake	Azure Databricks SQL Serverless
Opção de preço de computação/serviço	Edição Enterprise com compromisso de um ano (pago mensalmente)	Preço inicial total da instância reservada por um ano (34% de economia) (pagamento antecipado anual)	Edição Enterprise (AWS Leste dos EUA) (US\$ 3,00/crédito)	Unidades DBCU pré-adquiridas do Azure Databricks SQL Serverless com 1 ano de antecedência para 23% de desconto. (Nível Premium)
Configuração de computação/serviço	800 slots de capacidade comprometida. Escalonamento automático para 1.000 slots adicionais por uma média de 1 hora/dia e para 2.000 slots por uma média de 5 minutos/dia	Instâncias grandes 8 x ra3,16x	18 créditos (24 x 7) 22 créditos (1,2 hora/dia) 44 créditos (0,26 hora/dia)	Instância extragrande (80 DBU) (24 x 7) Escalonamento automático para picos: 2x cópias horizontais (média de 1,2 hora por dia) 3x cópias horizontais (média de 0,26 hora por dia)
Custo anual de capital para gastos iniciais	N/A	8%	8% (somente storage)	8%
Storage na nuvem	99,56 TB de storage físico (compactado) (otimizado automaticamente para economia de custos)	99,56 TB de storage gerenciado	99,56 TB de storage de capacidade (pago antecipadamente por 43% de economia)	99,56 TB de storage Blob Azure
Serviço de transmissão/carregamento de dados	Inserções de transmissão	AWS Kinesis	Snowpipe estimado como um grande data warehouse (8 créditos x 8 horas/dia)	Configuração do Azure Databricks para ingestão e transmissão (nível padrão, 6 DBU)

Fonte: Enterprise Strategy Group, uma divisão da TechTarget, Inc.

Como isso não foi projetado como um estudo para mover cargas de trabalho locais para a nuvem, o ESG não estimou o tempo e o custo da migração inicial da solução e a complexidade para cada fornecedor nessa análise.³ Presumimos que seja igual para todas as soluções. O ESG também assumiu que o desempenho fornecido por essas configurações atenderia aos SLAs dos negócios, e nenhuma vantagem de desempenho direto foi dada a uma solução em relação à outra. Na prática, esses fatores podem variar muito com base no ambiente de uma organização e devem ser cuidadosamente considerados por ela. Na verdade, o BigQuery pode ter vantagens em ambos os aspectos, mas de forma conservadora não atribuímos nenhuma vantagem de custo a nenhuma solução com base no tempo e na complexidade para migrar ou no desempenho melhorado/mais previsível.

Com base em nossas suposições conservadoras, o ESG descobriu que a solução BigQuery proporcionou um custo total da propriedade (TCO) 54% menor do que o Azure Databricks SQL Serverless, um TCO 47% menor do que o AWS Redshift e um TCO 26% menor do que o Snowflake durante o período modelado de três anos. A Figura 5 mostra o TCO estimado de três anos para cada uma das quatro soluções de EDW baseadas na nuvem para atender às necessidades do cenário modelado.

Figura 5. Estimativa de três anos de custo total da propriedade da solução de data warehouse baseada em nuvem



Fonte: Enterprise Strategy Group, uma divisão da TechTarget, Inc.

Gastos iniciais com nuvem (investimento de capital pago antecipadamente)

Ao estimar os gastos com nuvem, comparamos de forma conservadora os preços da Edição Enterprise do Google Cloud com uma solução econômica para cada uma das três soluções concorrentes durante o período de três anos. Para o AWS Redshift, isso envolveu a assinatura de contratos, e o pagamento antecipado com um ano de antecedência para atingir o nível de economia informado (34% de economia) para recursos de computação. Para a computação do Azure Databricks SQL Serverless, comprar créditos de DBCU com um ano de antecedência gerou uma economia de 23%. Para quaisquer pagamentos feitos com um ano de antecedência, calculamos o

³ O ESG estudou o custo de mover cargas de trabalho locais para a nuvem em [The Economic Advantages of Migrating Enterprise Data Warehouse Workloads to Google BigQuery](#), março de 2019.

impacto financeiro do pagamento do serviço antes de o benefício ser recebido calculando um custo de capital conservador de 8% para qualquer parte do pagamento restante em que o benefício ainda não tenha sido realizado (como um cálculo decrescente de pagamento de juros). Deve-se observar, no entanto, que ao reservar a computação com antecedência, a solução se torna muito menos flexível, tornando mais difícil aproveitar as tecnologias aprimoradas ou dimensionar a implantação para reduzir os custos. Os descontos para essas soluções podem aumentar ainda mais reservando recursos ou créditos de pré-compra com até três anos de antecedência. No entanto, isso reduz ainda mais a flexibilidade da solução. Para nosso cenário, presumimos que a computação do BigQuery e do Snowflake não precisava de pagamento antecipado.

Custos de nuvem sob demanda

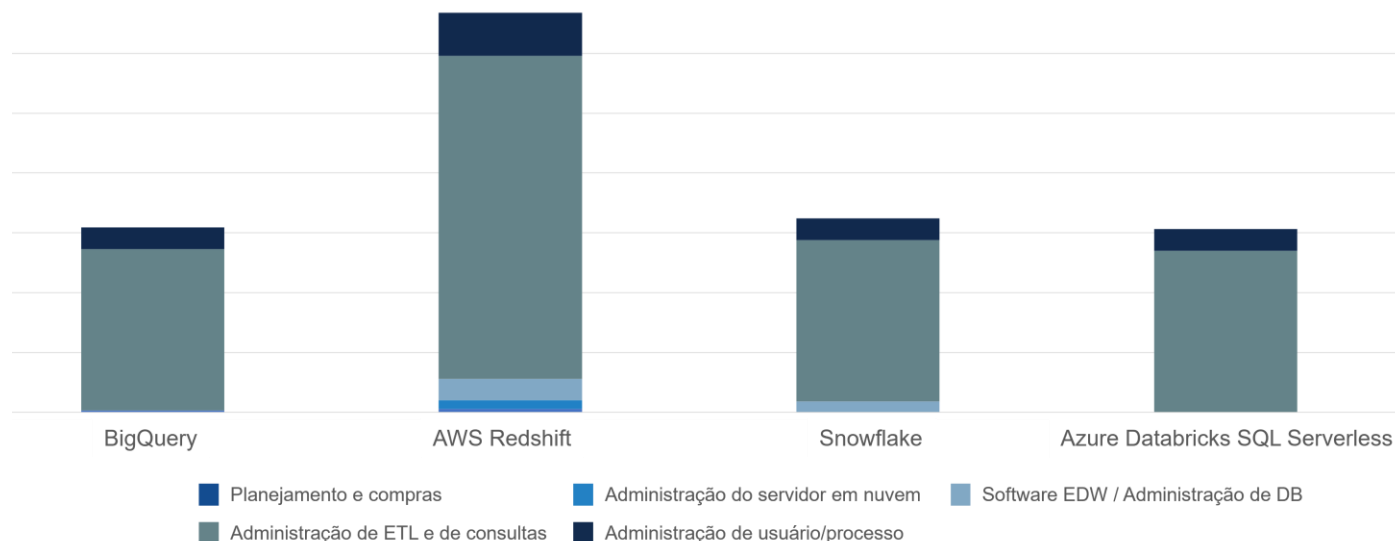
Os gastos mensais com nuvem consistiam na estimativa de faturamento que seria recebida pela organização a cada mês para computação, storage e/ou serviços. O custo de computação das instâncias do AWS Redshift RA3 foi pago antecipadamente anualmente ("Custos iniciais" na Figura 5). Os gastos mensais com nuvem ("Custos de nuvem sob demanda" na Figura 5) para o AWS Redshift consistiram em 99,56 TB de capacidade de storage gerenciado e o custo estimado do serviço de transmissão Kinesis para lidar com 760 GB de dados de transmissão. Todos os créditos de DBCU modelados para computação do Azure Databricks SQL Serverless também foram adquiridos anualmente. Os gastos mensais com nuvem do Azure Databricks SQL Serverless também consistiram em 99,6 TB de storage do Azure Blob e uma modesta configuração do Azure Databricks (6 DBU) para lidar com operações de ETL e transmissão. Os gastos mensais do Snowflake consistiam em um estado constante de 427 créditos diários para armazéns virtuais (18 créditos x 12 horas por dia), 38 créditos diários para lidar com picos de escalonamento automático, 64 créditos diários (8 créditos x 8 horas por dia) para o armazém virtual de carregamento de dados e o serviço Snowpipe e 99,6 TB de capacidade de storage. Os custos do BigQuery incluíram preços modelados para 800 slots de capacidade comprometidos, até 2.000 slots de escalonamento automático, custos físicos de storage na nuvem otimizados (dados com mais de 90 dias movidos para um storage mais econômico) e inserções de transmissão.

Custos administrativos

Os custos administrativos de cada uma das soluções foram estabelecidos com base em estimativas de boa-fé baseadas em opiniões de especialistas e usuários finais e em análises de estudos de caso. O ESG modelou as horas semanais ou ocasionais esperadas que um funcionário gastaria no planejamento e na compra, na administração de servidores virtuais ou nós em nuvem, na administração de software e banco de dados do EDW, na administração de ETL e consultas e na administração de usuários e processos. Todas as soluções baseadas em nuvem proporcionaram economias administrativas significativas em comparação com os modelos locais do ESG. O planejamento e a compra de todas as quatro soluções são muito simplificados em comparação ao planejamento de implantações no local, mas o AWS Redshift exigiria algum tempo gasto no dimensionamento e análise da solução antes de fazer o investimento inicial anual, enquanto o dimensionamento não ideal do ambiente do BigQuery, Snowflake ou do Azure Databricks SQL Serverless pode ser facilmente corrigido a qualquer momento.

Uma vez implantada, a solução Redshift exigiria monitoramento, configuração e ajuste de recursos virtualizados (instâncias ou nós), enquanto as outras soluções são totalmente gerenciadas nos bastidores. O BigQuery foi a única solução que não exigia ajuste e configuração do software e do banco de dados do EDW. O BigQuery, o Snowflake e o Azure Databricks SQL Serverless foram supostamente mais fáceis de gerenciar a partir de uma perspectiva de ETL, administração de consultas e administração de usuários/processos, e isso proporcionou a maior parte das economias administrativas em nossa análise. A ingestão simplificada do BigQuery e a capacidade de dar suporte a IA/ML nativo e integrar-se nativamente a outras ferramentas proporcionaram economias administrativas adicionais. A Figura 6 compara os custos administrativos modelados de três anos para cada uma das quatro soluções.

Figura 6. Estimativa de gastos administrativos de três anos para o cenário modelado do ESG



Fonte: Enterprise Strategy Group, uma divisão da TechTarget, Inc.

Questões a considerar

Os modelos do Enterprise Strategy Group são criados de boa-fé com base em suposições conservadoras, críveis e validadas; no entanto, nenhum cenário modelado jamais representará todos os ambientes possíveis. Os benefícios recebidos por uma organização dependem do tamanho da organização, da natureza do negócio e das capacidades do produto atual ou serviço em uso, além de muitas outras variáveis. Basear suposições nos requisitos de carga de trabalho em outros casos de uso pode não resultar no mesmo nível de economia que mostramos. O ESG recomenda que você realize sua própria análise de EDW disponível baseado em nuvem por meio de testes de prova de conceito.

Conclusão

Deve ser uma prioridade máxima de cada organização produzir inteligência de negócios de qualidade e insights práticos da maneira mais rápida e econômica possível. Os serviços de data warehouse baseados na nuvem oferecem às organizações de TI vantagens significativas em termos de agilidade e disponibilidade, ao mesmo tempo em que reduzem bastante o tempo de insight. Se uma organização pretende criar uma solução de análise de múltiplas nuvens, empregar uma estratégia de nuvem de dados unificada moderna, migrar sua operação de data warehouse no local para a nuvem, ou aumentar as operações existentes com serviços de EDW baseados na nuvem para lidar com novas oportunidades, é importante entender as diferenças entre as ofertas e escolher a solução que oferece à organização a maior agilidade, flexibilidade e interoperabilidade com outros serviços pelo menor custo.

O Enterprise Strategy Group (ESG) comparou o custo e os recursos das quatro principais soluções de EDW baseadas na nuvem: BigQuery, AWS Redshift, Snowflake e Azure Databricks SQL Serverless. O ESG usou entrevistas com usuários finais que tinham experiência com várias soluções, bem como estudos de caso para formar suposições conservadoras a fim de modelar o TCO necessário para atender às necessidades dos requisitos de EDW de uma organização modelada durante um período de três anos. Nosso modelo previu que o BigQuery poderia fornecer um TCO até 54% menor em uma solução que forneça operação mais simples e maior agilidade e escalabilidade, com recursos nativos e interoperabilidade com outros serviços importantes baseados na nuvem.

O BigQuery foi projetado para remover toda a carga física e lógica de gerenciamento, monitoramento, manutenção e proteção da infraestrutura de EDW, permitindo que os recursos organizacionais sejam concentrados com base na inteligência em vez de manter o processo de obtenção. O BigQuery é a única solução totalmente gerenciada (sem nós físicos ou virtuais para gerenciar e manter) com storage auto-otimizado e suporte nativo para cargas de trabalho de IA/ML. O BigQuery oferece economia em relação às outras soluções de EDW baseadas na nuvem, oferecendo operações mais simples e administração diária, eliminando a necessidade de dimensionar a implantação, gerenciar o ambiente e planejar o crescimento de recursos virtuais. Além disso, o BigQuery não requer investimento inicial e pode ser facilmente dimensionado para atender às necessidades em constante mudança dos negócios.

O BigQuery é a única solução de nuvem nativa que foi projetada desde o primeiro dia com recursos realmente sem servidor e de transmissão e tem fornecido e aperfeiçoado as ferramentas e os processos ao longo de 10 anos de implantações de clientes. Outros data warehouses na nuvem podem ter sido criados com o aproveitamento de projetos e princípios transmitidos por soluções locais, mas eles estão lentamente começando a perceber a importância dos recursos sem servidor e de transmissão e estão começando a implementá-los usando a automação integrada.

As soluções de EDW atuais devem atuar como um repositório global de informações, propiciar a agilidade para aumentar ou diminuir a demanda e integrar-se perfeitamente a outras funções analíticas e operacionais. Mas simplesmente comparar as tabelas de preços públicas não é suficiente; as organizações também devem considerar o custo das pessoas e do processo. O ESG recomenda que todas as organizações realizem sua própria análise para comparar soluções, e esperamos que esse relatório ajude a identificar as áreas que devem ser consideradas ao comparar serviços de EDW baseados na nuvem.

Com as organizações lançando um olhar cada vez atento sobre a otimização de custos, deve-se observar que o escalonamento automático refinado do BigQuery permite que as organizações operem de forma previsível com capacidade de 30% a 40% menos comprometida, e a flexibilidade de escolha para opções de preços de computação e storage permitirá que as organizações controlem melhor seus gastos com nuvem para cada projeto ou carga de trabalho. As recentes mudanças no modelo de cálculo e de preço de storage beneficiaram os clientes do BigQuery com os quais conversamos. Como disse um desses clientes: "As mudanças no modelo de preços do BigQuery não só nos pouparam dinheiro, mas também trouxeram mais consistência ao nosso faturamento, que costumava variar muito dependendo do nosso uso." Recomendamos fortemente que as organizações considerem o BigQuery em conjunto com outros serviços baseados em nuvem do Google Cloud para potencializar a plataforma que fornece insights de negócios de última geração para toda a organização.

©TechTarget, Inc. ou suas subsidiárias. Todos os direitos reservados. TechTarget e o logotipo TechTarget são marcas comerciais ou marcas registradas da TechTarget, Inc. e estão registradas em jurisdições em todo o mundo. Outros nomes e logotipos de produtos e serviços, inclusive BrightTALK, Xtelligent e Enterprise Strategy Group, podem ser marcas registradas da TechTarget ou de suas subsidiárias. Todas as outras marcas registradas, logotipos e nomes de marcas são propriedade de seus respectivos donos.


As informações contidas nesta publicação foram obtidas por fontes que a TechTarget considera confiáveis, mas não são garantidas pela TechTarget. Esta publicação pode conter opiniões da TechTarget, que estão sujeitas a alterações. Esta publicação pode incluir previsões, projeções e outras declarações preditivas que representam as suposições e expectativas da TechTarget, Inc. à luz das informações atualmente disponíveis. Essas previsões são baseadas nas tendências do setor e envolvem variáveis e incertezas. Consequentemente, a TechTarget não oferece nenhuma garantia quanto à precisão das previsões, projeções ou declarações preditivas específicas aqui contidas.

Qualquer reprodução ou redistribuição desta publicação, integral ou parcialmente, seja em formato impresso, eletrônico ou outro, para pessoas não autorizadas a recebê-la, sem o consentimento expresso da TechTarget, viola a lei de direitos autorais dos EUA e estará sujeita a ação de indenização cível e, se for o caso, criminal. Em caso de dúvidas, entre em contato com o setor de Relações com clientes em cr@esg-global.com.

Sobre o Enterprise Strategy Group

O Enterprise Strategy Group da TechTarget fornece inteligência de mercado focada e acionável, pesquisa de demanda, serviços de consultoria de analistas, orientação estratégica de GTM, validações de soluções e conteúdo personalizado que dá suporte à compra e venda de tecnologia empresarial.

 contact@esg-global.com

 www.esg-global.com