



TURBINE SEU DATA LAKE COM O SNOWFLAKE

DATA LAKES: A PROMESSA

Há dez anos, os data warehouses herdados não conseguiam acompanhar os volumes crescentes e a grande variedade de dados coletados por empresas e que precisavam de análise: desde fontes tradicionais, como sistemas ERP e CRM, a dados de fontes de terceiros e fontes mais modernas, como blogs, dispositivos de Internet das Coisas e redes sociais. A princípio, a solução para gerenciar grandes volumes de dados parecia ser o data lake, onde as organizações poderiam coletar, armazenar e analisar todos os seus dados estruturados e não estruturados em um só lugar. Começando por soluções locais e então partindo para a nuvem, esperava-se que os data lakes fossem capazes de solucionar o problema com custo e complexidade mínimos.

A promessa inicial do data lake era:

- Aproveitar a variedade, a velocidade e o grande volume de dados
- Armazenar muitos tipos de dados em seus formatos nativos e disponibilizá-los à comunidade empresarial para geração de relatórios e fins de ciência de dados
- Permitir às organizações explorar, refinar e analisar petabytes de informação recebidas continuamente de diversas fontes de dados

Essa promessa não se cumpriu e, até o momento, o retorno do investimento em data lakes tem sido baixo.

POR QUE AS ORGANIZAÇÕES NÃO OBTIVERAM VALOR

Muitos data lakes se tornaram aterros virtuais de dados que dificilmente renderão análises úteis. A Figura 1 resume os desafios inerentes a muitos data lakes:

- Dados em silos, sem uso, com desempenho de consulta limitado
- Pipelines de dados complexos e degradação de desempenho
- Dados não seguros e não governados

DADOS INACESSÍVEIS E DESEMPENHO DE CONSULTA LIMITADO

Colocar todos os tipos de dados brutos em um data lake é relativamente simples, mas obter insights com base neles é bem mais desafiador. Poucos técnicos conseguem fazer isso, o que exclui a maioria dos tomadores de decisões de negócios que precisam de análises atualizadas. Para ser usado por completo, um data lake deve democratizar o acesso aos dados: pessoas de todos os níveis de habilidade (incluindo gerentes de primeira linha, analistas financeiros, executivos, cientistas de dados e especialistas de engenharias de dados) devem ter acesso fácil e apropriado ao sistema para realizarem as análises de que precisam.

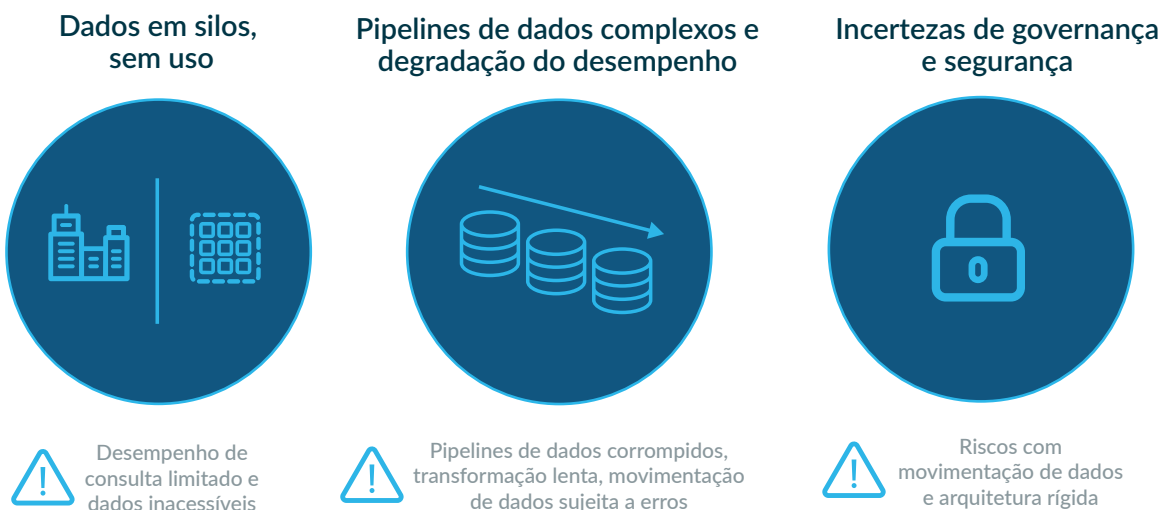


Figura 1: os data lakes possuem desafios inerentes.

Infelizmente, os data lakes podem facilmente se tornar repositórios infinitos de dados, desorganizados e sem estrutura. Assim, eles tendem a ser subutilizados e o valor dos dados não é totalmente aproveitado.


Além disso, o desempenho de consulta muitas vezes é limitado devido a problemas de simultaneidade e escalabilidade. Por isso, os analistas muitas vezes acabam com um desempenho de análise de dados limitado e grande parte dos dados inacessíveis no data lake.

PIPELINES DE DADOS CORROMPIDOS E TRANSFORMAÇÃO LENTA

A maioria dos dados precisa ser transformada antes de ser consumida. Mas o processo de transformação geralmente envolve uma combinação de diversas opções de ferramentas e programação para atender aos diferentes usuários e casos de uso. A arquitetura muitas vezes é complexa, com movimentos de dados sujeitos a erros fora do data lake e muitas tecnologias alinhadas. Tudo isso contribui para que pipelines de dados rompam sob cargas de trabalho pesadas e o desempenho caia. Os engenheiros de dados também despendem tempo construindo pipelines complexos, configurando e gerenciando infraestruturas complexas, com muitos “botões para apertar”.

DADOS NÃO SEGUROS E NÃO GOVERNADOS

Com regulamentações como a GDPR e demais normas regionais de privacidade em vigor, as empresas não têm escolha senão governar, monitorar e entender seus dados com precisão: sua qualidade, os casos de uso, a trilha de auditoria e a acessibilidade. A implementação, porém, é desafiadora. As organizações precisaram fazer uma escolha difícil: empregar uma arquitetura de dados rígida com regras estritas de segurança e governança ou afrouxar o cinto para que os usuários pudessem usar os dados para desempenhar suas tarefas. Tem sido extremamente desafiador para as organizações fornecer uma arquitetura flexível com segurança e governança adequadas que permita aos usuários usar os dados de que precisam.



**“Ao longo de 2022,
apenas 20% dos
insights analíticos
forneceram resultados
de negócios.”**

Andrew White

VP e analista de renome da
Gartner
Janeiro de 2019

ARQUITETURAS DE DADOS COMPLEXAS

As arquiteturas de dados tradicionais não foram criadas para atender aos requisitos de uma análise de dados moderna e uma engenharia de dados eficiente. Dados em formatos diferentes muitas vezes chegam a sistemas diferentes sem um jeito simples de unificar, o desempenho e a confiabilidade do pipeline são afetados pela contenção de recursos, sendo demorado construir e manter um pipeline de dados com uma infraestrutura de gerenciamento complexa.

A Figura 2 mostra uma arquitetura de dados típica de muitas empresas: uma rede complexa de tecnologias combinadas que leva os dados desde a origem até o destino desejado. Ao longo das etapas de integração, transformação, agregação e entrega, as organizações implementam múltiplas ferramentas,

sendo que cada ponto de conexão é uma falha em potencial. Problemas de governança, segurança e isolamento são comuns nesse tipo de abordagem, com várias cópias dos dados surgindo para atender às necessidades variadas das diferentes áreas de um negócio.

No final das contas, as organizações passam mais tempo gerenciando a infraestrutura e menos tempo trabalhando, de fato, com os dados. As equipes de TI ficam sobrecarregadas com uma infinidade de problemas técnicos, incapazes de oferecer uma resposta ágil às solicitações de negócios. Analistas e tomadores de decisões ficam limitados a dados incompletos e obsoletos, gerando incerteza nos resultados das análises.

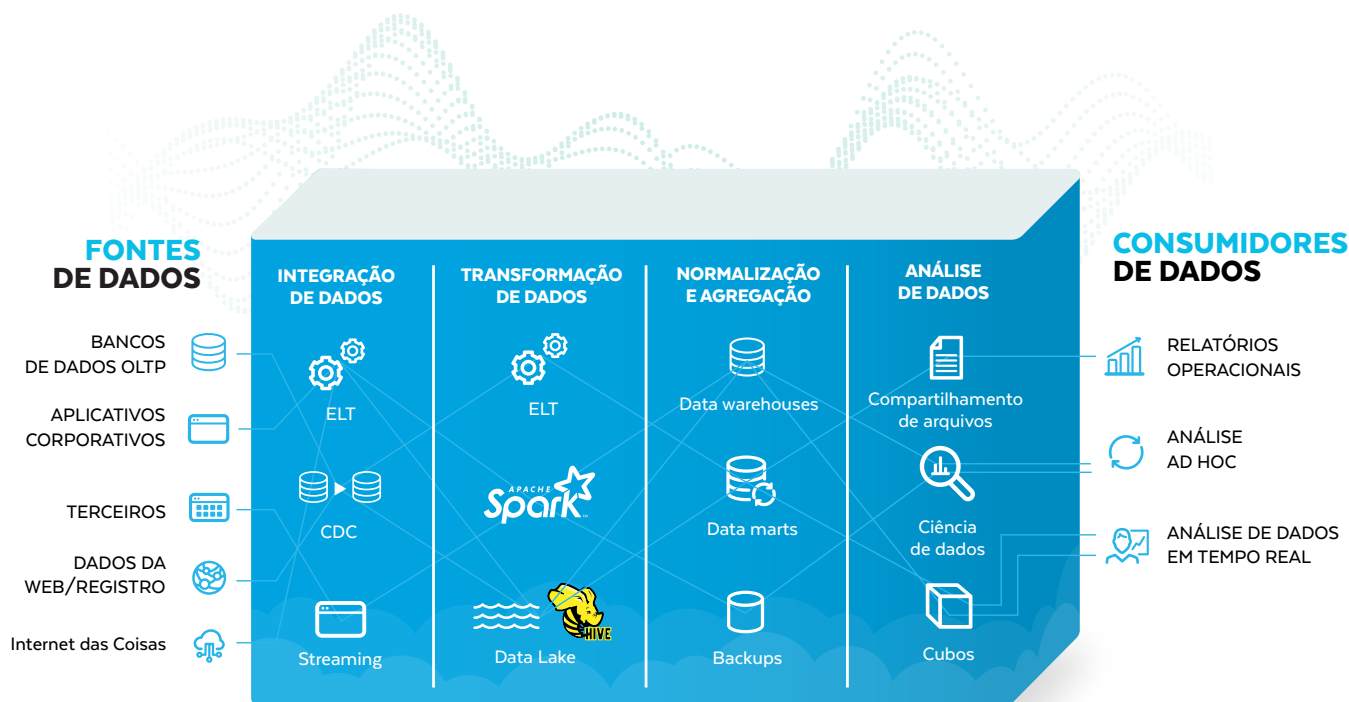


Figura 2: a arquitetura de dados tradicional é complexa, dispendiosa e restrita.

MANTENHA SEU DATA LAKE, MAS TURBINE-O COM O SNOWFLAKE

Sem importar os seus requisitos de desempenho de consulta, a natureza dos seus pipelines de dados ou as exigências de segurança e governança, a plataforma Snowflake é flexível o bastante para atender às necessidades das suas cargas de trabalho e usuários. Caso você já tenha investido em um data lake na nuvem e planeje mantê-lo, é possível usar o Snowflake para aprimorar e complementar o data lake existente a fim de obter um retorno rápido do investimento. Além disso, é possível:

- Facilitar o acesso a dados usando o Snowflake como um mecanismo eficiente de consulta junto ao seu data lake, com desempenho excelente sem problemas de contenção de recursos ou simultaneidade.
- Usar o Snowflake para executar transformações de dados confiáveis e de alto desempenho, com pipelines de dados integrados extensíveis para descarregar os dados com facilidade de volta ao seu data lake.
- Possibilitar maior colaboração com os seus dados, aproveitando a segurança e a governança integradas na plataforma Snowflake.

USO DO SNOWFLAKE COMO UMA CAMADA DE CONSULTA RÁPIDA

Um jeito simples de obter um retorno rápido do investimento no seu data lake na nuvem é usar o Snowflake como uma camada de consulta rápida junto a ele. Ao usar o Snowflake para consultar dados onde eles residem (ou seja, em armazenamentos de nuvem como Amazon S3, Microsoft Azure Blob Storage ou Google Cloud Storage), é possível manter o data lake como uma única fonte de verdade, mesmo em ambientes multinuvm. Uma única fonte de verdade elimina a tarefa árdua de manter diversos repositórios de dados em sincronia, facilitando o acesso aos dados para os tomadores de decisão.

Como mostrado na Figura 3 e na tabela da página seguinte, bem como nas próximas seções, o Snowflake oferece vários métodos para consultar e interpretar os dados em seu data lake, mantê-lo como única fonte de verdade e liberar o potencial total de seus dados.

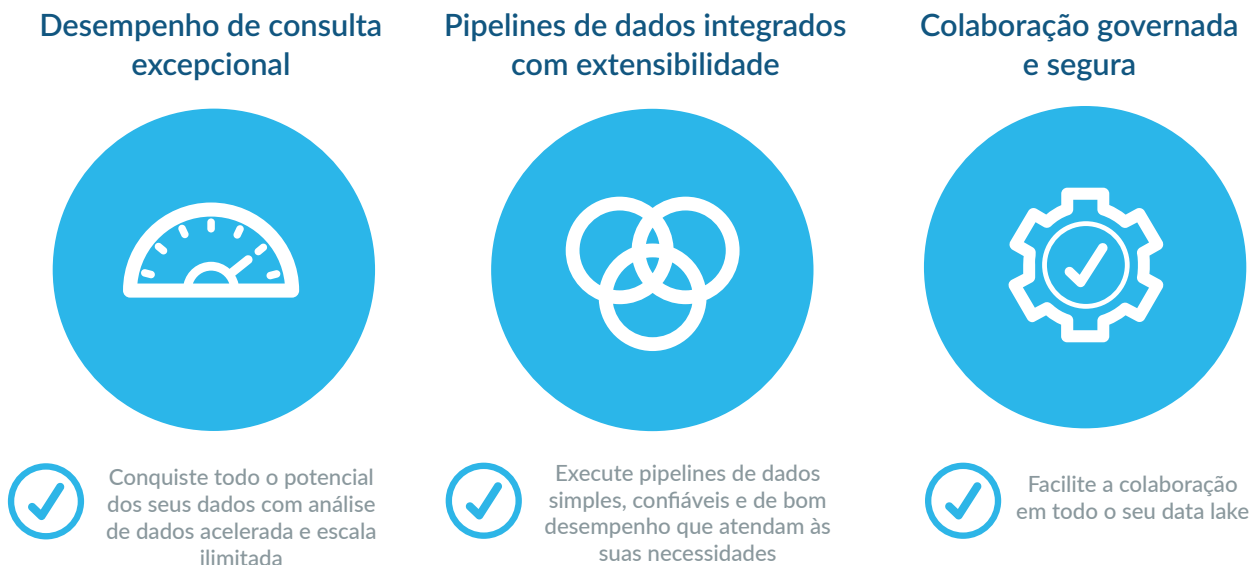


Figura 3: o Snowflake maximiza a utilidade do seu data lake pré-existente.

USE ISSO:	PARA FAZER ISSO:
Tabelas externas	Consultar dados diretamente no seu data lake
Serviço de notificação da sua nuvem	Atualizar metadados de tabela externa automaticamente
Visualizações materializadas das suas tabelas externas	Acelerar consultas frequentes
Atualização automática de partições	Registro automático de arquivos novos
Snowsight (integrado na interface do Snowflake)	Acelerar a exploração de dados com tabelas e painéis integrados

TABELAS EXTERNAS CONSULTAM DIRETAMENTE O DATA LAKE

As tabelas externas são usadas somente para leitura sobre os arquivos no seu armazenamento de nuvem externo. Use-as para consultar dados diretamente no seu data lake sem movê-los para o Snowflake. As tabelas externas armazenam metadados sobre o arquivo em seu nível, incluindo o caminho do arquivo, um identificador de versão e informação sobre o particionamento. Elas possibilitam consultar os dados armazenados nos arquivos do seu data lake como se tudo fizesse parte de um único banco de dados.

As tabelas externas são somente leitura. Então, não é possível fazer operações de DML nelas. No entanto, é possível usá-las em operações de consulta e junção. Também é possível criar visualizações de tabelas externas.

Como elas armazenam metadados no nível do arquivo sobre os arquivos de dados, os usuários podem consultar os dados armazenados nos arquivos como se estivessem lidando com uma tabela que de fato contém os dados. Tudo isso pode ser extremamente útil quando a situação requer uso analítico somente para leitura dos dados armazenados em um local. Em vez de carregar o conjunto de dados inteiro em uma tabela que ocupa espaço no banco de dados, o espaço externo permite que os usuários façam transformações, cálculos e outras análises.

Com as tabelas externas é possível:

- Consultar dados diretamente no Amazon S3, Microsoft Azure Blob Storage ou Google Cloud Storage.
- Manter seu data lake como única fonte de verdade, eliminando a necessidade de copiar ou transferir dados.
- Obter uma análise rápida dos dados de uma fonte externa, onde quer que estejam.

A ATUALIZAÇÃO AUTOMÁTICA DAS TABELAS EXTERNAS MANTÉM TUDO ATUALIZADO

O serviço de notificação da sua nuvem alerta o Snowflake sempre que os metadados da tabela externa precisam ser atualizados, e o Snowflake o faz automaticamente. O serviço de notificação alerta o Snowflake quando:

- Novos arquivos no local forem adicionados aos metadados da tabela.
- Mudanças nos arquivos no local forem atualizadas nos metadados da tabela.
- Arquivos que não estão mais no local forem removidos dos metadados da tabela.

Os dados externos estão sempre atualizados e disponíveis aos usuários para consulta e acesso imediato, como mostrado na Figura 4 da página seguinte.

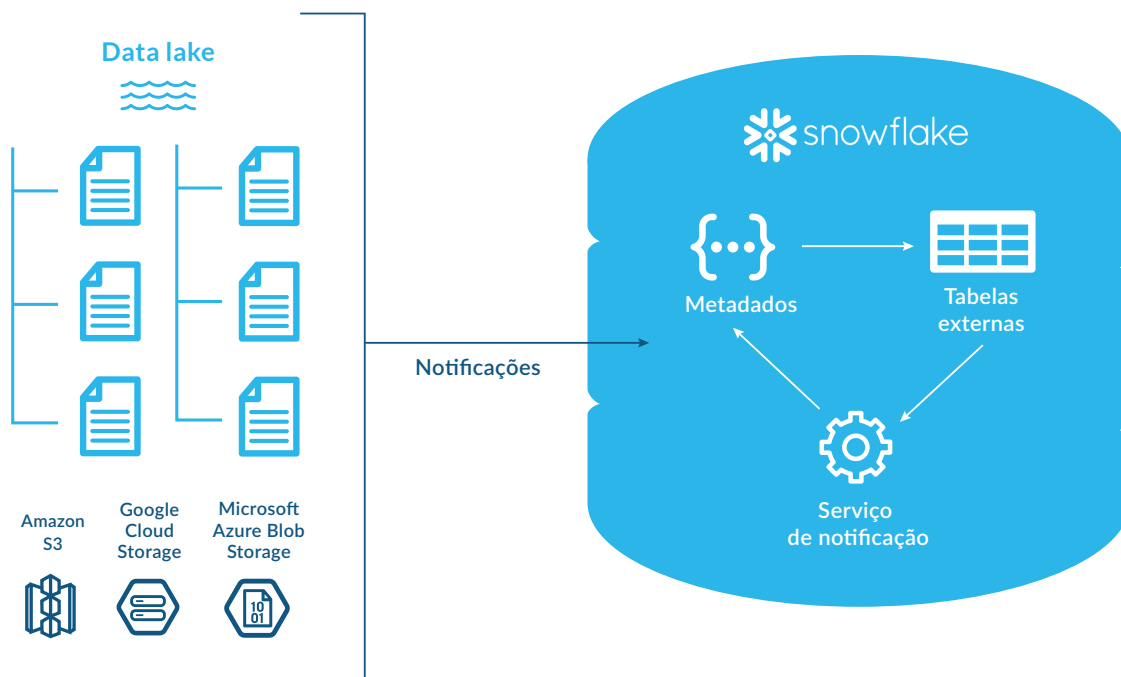


Figura 4: o Snowflake recebe alertas quando partições da tabela externa mudarem, atualizando os metadados da tabela automaticamente.

VISUALIZAÇÕES MATERIALIZADAS ACELERAM CONSULTAS FREQUENTES

Consultar dados armazenados fora do banco de dados talvez demore mais do que consultar tabelas nativas do Snowflake. No entanto, criar visualizações materializadas sobre as tabelas externas pode melhorar significativamente o desempenho das consultas. Uma visualização materializada é um objeto de banco de dados que armazena seus próprios dados no Snowflake (como uma tabela). Diferente de uma visualização, não se trata de uma janela para o armazenamento de outro objeto, mas sim um objeto separado que contém seus próprios dados, atualizados periodicamente. As visualizações materializadas aceleram resultados de consultas em conjuntos de dados volumosos, diminuindo o impacto nos seus recursos de rede.

Consultar uma visualização materializada (sobre uma tabela externa) oferece um desempenho mais rápido do que uma consulta à tabela externa subjacente. A diferença no desempenho pode ser significativa quando a consulta é feita com frequência ou é complexa.

As visualizações materializadas podem melhorar o desempenho de consultas que empregam repetidamente os mesmos resultados de subconsultas. O Snowflake mantém essas visualizações de forma automática e transparente. Um serviço em segundo plano atualiza a visualização materializada após ocorrerem mudanças na tabela base. O processo é mais eficiente e menos sujeito a erros do que a manutenção manual do equivalente de uma visualização materializada no nível da aplicação.

Os dados acessados por meio de visualizações materializadas estão sempre atualizados, não importando quantas operações de DML foram conduzidas na tabela base. Se uma consulta é executada antes da visualização materializada ser atualizada, o Snowflake a atualiza ou usa partes atualizadas dela para recuperar dados novos necessários da tabela base.

As visualizações materializadas são especialmente úteis quando:

- As consultas possuem filtros complexos ou dinâmicos e predicados de junção que reduzem o número de linhas a serem examinadas.
- As consultas são muito seletivas em relação ao número de colunas necessárias para satisfazer um resultado de consulta.
- Os resultados de consulta exigem processamento inteligente, incluindo:
 - Análise de dados semiestruturados
 - Agregados que demoram para ser calculados
- A consulta ocorre em uma tabela externa (ou seja, conjuntos de dados armazenados em arquivos em um espaço externo), que pode ter um desempenho mais lento em relação às consultas de tabelas de banco de dados nativas.
- As consultas à tabela base ou externa possuem agregações ou predicados de filtro comuns e não atendem aos requisitos de desempenho dos usuários finais.
- A tabela base da visualização tem baixa rotatividade.

Uma visualização materializada é criada nas condições a seguir:

- O pré-processamento de dados (agregação, criação de clusters, ingestão de uma tabela externa e assim por diante) contribui para melhoras significativas de desempenho nas consultas executadas por usuários finais.
- As visualizações materializadas são consultadas e usadas com frequência.
- A economia de consultas com melhor desempenho (warehouses menores para atender aos requisitos de SLA) superam os custos de pré-processamento e manutenção de visualizações materializadas.

SNOWSIGHT AGILIZA A EXPLORAÇÃO DE DADOS

A interface do Snowflake conta com uma ferramenta de exploração de dados com base em SQL feita para usuários que querem visualizar os resultados de consultas sem o suporte de um engenheiro de dados. O Snowsight possibilita a colaboração em toda a sua organização. É possível compartilhar consultas e painéis com outros membros da sua organização enquanto se mantém o controle com permissões granulares. O Snowsight também oferece visualizações detalhadas integradas no Snowflake para que você passe de dados a insights mais rapidamente.

O Snowsight ainda inclui benefícios e recursos como:

- Consultas rápidas e com capacidade de resposta: o Snowsight conta com um editor ágil de nível de desktop.
- Preenchimento automático inteligente: o Snowsight faz sugestões contextuais baseadas na sua consulta e dialeto SQL.
- Resultados interativos: é possível pré-visualizar dados rapidamente, não importando quantas linhas uma consulta retorna.
- Gráficos visualmente melhores: a boa visualização dos gráficos é mantida em diversos dispositivos com qualquer tamanho.
- Painéis modernos: o Snowsight oferece uma interface arraste-e-solte simples para a criação de painéis.
- Filtros dinâmicos de dados: use parâmetros para definir datas, clientes e muito mais.
- Links privados compartilháveis: envie aos seus colegas consultas ou painéis com um link para que eles possam visualizar, executar ou editar o conteúdo.

COMBINANDO TUDO

Como mostrado na Figura 5, o mecanismo de consulta do Snowflake complementa seu data lake atual e permite acessar os dados armazenados de maneira rápida e fácil. Consulte o data lake por meio de tabelas externas atualizadas automaticamente quando há mudança nos metadados, use visualizações materializadas para acelerar consultas frequentes e use o Snowsight, baseado em SQL, para interpretar visualmente os dados e compartilhá-los com colegas.

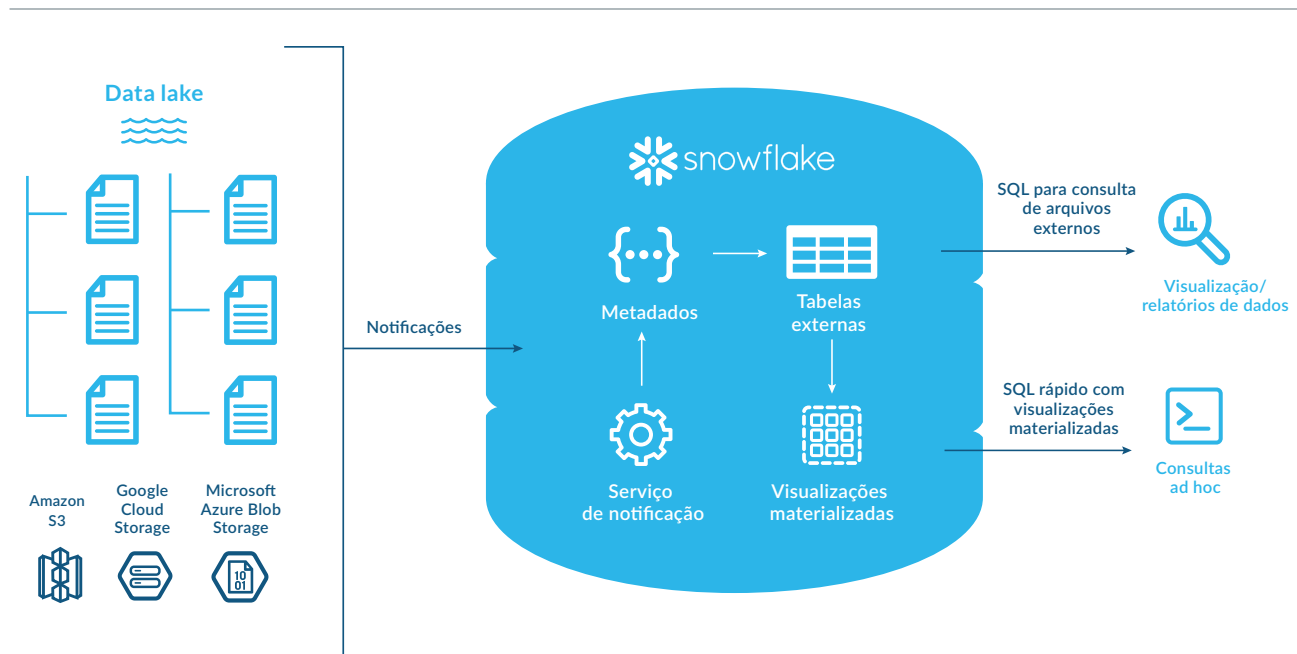


Figura 5: o uso do Snowflake como um mecanismo avançado de consultas permite uma rápida análise de dados e acesso a datas lakes existentes.

USE O SNOWFLAKE COMO UM MECANISMO AVANÇADO DE TRANSFORMAÇÃO

O armazenamento de objetos na nuvem se tornou a primeira opção para os data lakes. Embora os data lakes possam armazenar grandes volumes de dados, transformar e usar tudo isso para obter insights pode ser desafiador. O Snowflake permite que você obtenha um retorno rápido do investimento e atende às suas necessidades de transformação de dados com uma arquitetura simplificada que complementa seu data lake.

Dependendo da sua arquitetura, é possível usar tabelas externas para acessar seu data lake diretamente na nuvem. Também é possível usar o Snowpipe, o serviço de ingestão sem servidor do Snowflake, para ingerir dados automaticamente no Snowflake. Em seguida, a funcionalidade Streams e Tasks do Snowflake captura dados alterados e processa, continuamente, dados novos ou alterados. Após a transformação, é possível descarregar os dados de volta no seu data lake, particionados automaticamente pela coluna que você escolher, com a funcionalidade Data Lake Export do Snowflake (Consulte a Figura 6.)

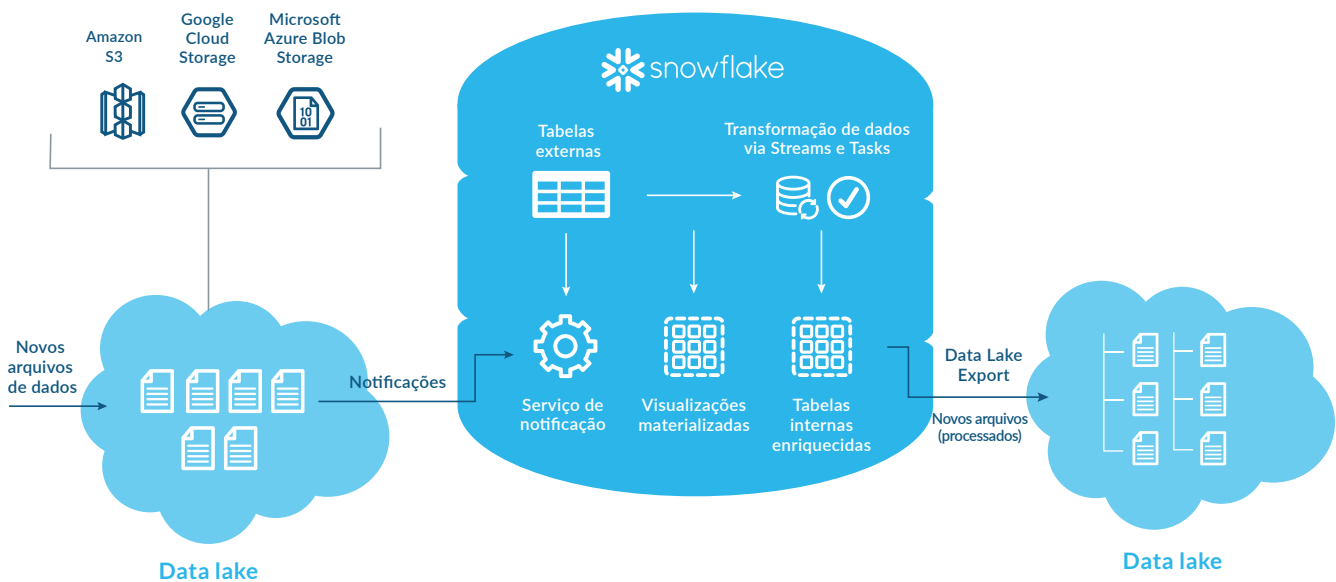


Figura 6: use o Snowflake como mecanismo de transformação para turbinar seu data lake.

PIPELINES CONTÍNUAS AUTOMATIZAM O FLUXO DE DADOS

O serviço de ingestão sem servidor do Snowpipe atua em segundo plano para carregar de forma assíncrona os dados do seu armazenamento em nuvem. A Figura 7, na página seguinte, detalha os benefícios de usar o Snowpipe para ingestão de dados.

RECURSOS DE CDC MONITORAM AS MUDANÇAS

Você pode criar pipelines de dados contínuos com o Snowpipe ou usar tabelas externas para apontar para os dados no seu armazenamento de nuvem, e então usar Streams e Tasks do Snowflake para agendar tarefas de ingestão de dados e capturar de dados alterados.

- Um Stream é um tipo de objeto do Snowflake com recursos de captura de dados alterados (Change Data Capture, CDC) para monitorar as mudanças realizadas em uma tabela, incluindo inserções e alterações em DML. Dessa forma, é possível agir com base nos dados alterados. Um Stream de tabela permite consultar uma tabela e consumir uma série de mudanças em uma tabela, no nível da linha, entre dois pontos transacionais no tempo.
- Uma Task é um tipo de objeto do Snowflake que define um cronograma recorrente para execução de instruções SQL, incluindo aquelas que chamam procedimentos armazenados. É possível combinar Tasks para execução sucessiva, dando suporte a processamento periódico de maior complexidade.

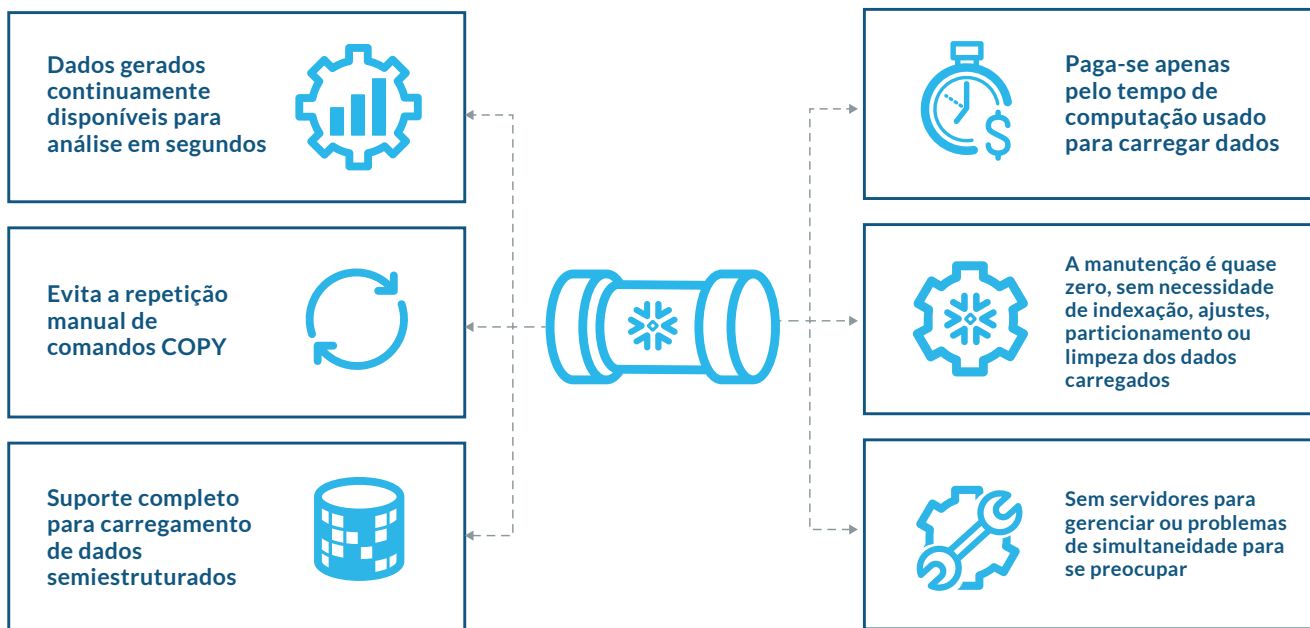


Figura 7: o Snowpipe simplifica a movimentação de arquivos de armazenamentos de objetos para o Snowflake.

Em um pipeline de dados contínuos, as Tasks podem usar Streams opcionalmente para maior conveniência no processamento contínuo de dados novos ou alterados. Uma Task pode verificar se um Stream contém dados alterados para uma tabela e consumir os dados alterados ou pular a operação atual, se não houver dados alterados.

Após uma transformação, é possível descarregar os dados transformados no seu data lake com o recurso Data Lake Export. As funcionalidades Streams e Tasks do Snowflake facilitam o agendamento de tarefas de integração de dados e a captura de dados alterados. Ou seja, não é preciso carregar os dados o tempo todo, sendo possível apenas processar os dados alterados.

PROCEDIMENTOS ARMAZENADOS AUTOMATIZAM TAREFAS FREQUENTES

Um procedimento automatiza uma tarefa executada com frequência que exige várias instruções SQL. É possível criar um procedimento uma vez e executá-lo quando necessário.

Para criar um procedimento armazenado, usa-se JavaScript e, na maioria das vezes, SQL: o JavaScript oferece as estruturas de controle (ramificações e loops), enquanto o SQL é executado ao chamar funções em um uma API JavaScript.

Os procedimentos armazenados oferecem:

- Lógica procedural (ramificações e loops), não suportada por SQL
- Tratamento de erros
- Capacidade de criar uma instrução SQL de forma dinâmica e executá-la
- Capacidade de criar um código que é executado com os privilégios da função proprietária do procedimento, em vez de com os privilégios da função que executa o procedimento

PIPELINES DE DADOS EXTENSÍVEIS POSSIBILITAM TRANSFORMAÇÕES ROBUSTAS DE DADOS

Muitas transformações de dados são extremamente complexas e difíceis de criar. Por exemplo, elas podem ser criadas com outras linguagens ou estruturas, ou ainda, usarem código de terceiros e serviços externos. No passado, seria preciso gerenciar diferentes serviços e sistemas, muitas vezes em ambientes distintos de dados com arquiteturas complexas. Já o Snowflake possibilita pipelines extensíveis em que é possível definir funções externas e também usar serviços de terceiros.

Os usuários podem criar e chamar seus próprios serviços remotos ou chamar serviços remotos criados por terceiros. Esses serviços remotos podem ser criados usando qualquer conjunto de servidores HTTP, incluindo serviços de computação em nuvem sem servidor, como AWS Lambda.

Alguns exemplos de funções externas:

- Usar serviço de geocodificação para aprimorar endereços com coordenadas e regiões políticas
- Usar serviço terceirizado para realizar análise de percepção de mensagens
- Conferir pontuações a clientes por meio de um modelo personalizado de aprendizado de máquina
- Usar lógica personalizada para extrair endereços de e-mail a partir de e-mails
- Obter preços de ações de um serviço remoto

ARQUITETURA COMPUTACIONAL MULTICLUSTER OFERECE ELASTICIDADE E ESCALA

Os pipelines de dados no Snowflake podem operar em lotes ou continuamente, e o processamento pode ocorrer dentro do próprio Snowflake. Graças à abordagem de computação multicluster do Snowflake, esses pipelines podem lidar com transformações complexas sem afetar o desempenho de outras cargas de trabalho.

Como isso funciona? Com a arquitetura de dados compartilhados multicluster do Snowflake, os recursos de computação e armazenamento são separados fisicamente, mas fazem parte de um único sistema de data warehouse integrado no nível lógico. A arquitetura multicluster exclusiva permite suporte a quantas cargas de trabalho isoladas forem necessárias. Cada carga de trabalho tem seu próprio mecanismo de computação e pode ser ampliada ou reduzida conforme necessário, sem a necessidade de alocar recursos previamente ou interromper outros processos.

Com essa arquitetura inovadora, o Snowflake pode suportar, com facilidade, várias cargas de trabalho diferentes. Por conta da separação entre computação e armazenamento, é possível ampliar armazenamentos virtuais de diferentes tamanhos para executar seus processos de ELT ou apoiar usuários de relatórios de inteligência de mercado, cientistas e mineradores de dados, sem contenção de recursos.

SUORTE NATIVO A DADOS SEMIESTRUTURADOS SIMPLIFICA A ARQUITETURA DE TRANSFORMAÇÃO

Formatos semiestruturados, como Parquet, ORC, JSON, XML e Avro não são fortemente tipificados. Geralmente se vê uma hierarquia de nós. Cada nó tem um nome e um valor, podendo ser individual e

constante, uma matriz de pares de nome/valor ou um objeto aninhado. Os bancos de dados tradicionais não têm sido capazes de lidar com tal estrutura variável, então os desenvolvedores recorrem ao uso de ETL, criando transformações complexas que nivelam os dados em uma estrutura tradicional de linhas e colunas.

O Snowflake introduziu um tipo especial para dados semiestruturados, chamado VARIANT. Veja um exemplo de como o processo de ingestão de dados semiestruturados funciona: o armazenamento virtual registra todo o documento JSON ou Parquet, em seu estado original, na coluna VARIANT. Se você consultar essa coluna no SQL, terá o mesmo documento JSON ou Parquet.

Ao mesmo tempo, a camada de serviços globais coleta metadados específicos dos conteúdos do documento, como nomes de nós, hierarquias e estruturas de matriz, e registra tudo isso no armazenamento de metadados. O Snowflake também organiza internamente os dados em colunas, permitindo a consulta direta no SQL. Esses dados em colunas são registrados em arquivos de micropartição, mas não se vê os nomes das colunas na definição da tabela (cabe notar que não há nomes de coluna em nenhum dos nós JSON ou Parquet). É possível referenciar os nós da coluna VARIANT com uma notação de ponto simples no SQL. O processo tem excelente desempenho e elimina a necessidade de realizar transformações explícitas.

USE O SNOWFLAKE PARA DADOS GOVERNADOS E SEGUROS

Não importando onde seus dados residem, o Snowflake garante governança e segurança globais. Diferente das soluções tradicionais que tornam difícil oferecer segurança granular, controle de acesso baseado em função (Role-Based Access Control, RBAC) ou compartilhamento seguro de dados, o Snowflake conta com fortes recursos de governança em sua plataforma, tais como:

- RBAC
- Segurança e controle de acesso no nível da linha por tabelas de titularidade
- Segurança no nível da coluna de dados externos com opções de mascaramento e tokenização
- Compartilhamento e troca de dados ativos seguros (como camada de consulta via tabelas externas)
- Segurança e governança para dados semiestruturados

Com essa abordagem integrada, sua equipe de segurança e governança pode:

- Definir controle de acesso granular para uma acessibilidade precisa
- Aproveitar funções de mascaramento no nível da coluna e filtragem no nível da linha para garantir segurança refinada
- Maximizar a capacidade para usuários trabalharem com dados sem revelar informações confidenciais
- Conduzir melhores análises de dados e colaboração por meio de compartilhamento seguro de dados ativos com usuários internos e externos

OU... SUBSTITUA SEU DATA LAKE PELO SNOWFLAKE

Caso atenda às suas necessidades de dados, você pode optar por usar o Snowflake como data lake e data warehouse. Com o Snowflake, é possível configurar um estágio de dados brutos e compactar seus armazenamentos de dados, de três a cinco vezes. Ao fazê-lo, você conta com os benefícios de escalabilidade, desempenho e facilidade de uso mencionados antes e também simplicidade: uma única plataforma e uma única cópia dos dados, suportando várias cargas de trabalho e casos de uso, com governança completa em diferentes nuvens.

Ao substituir sua infraestrutura atual de data lake pelo Snowflake, você também pode optar pelo Snowflake Data Cloud. O Data Cloud elimina o isolamento de conjuntos de dados em armazenamentos de nuvem e data centers locais, possibilitando que as organizações unifiquem e conectem-se a uma única cópia de todos seus dados com facilidade. O Data Cloud é um ecossistema de milhares de organizações conectando-se aos seus próprios dados e umas às outras ao compartilhar dados de maneira contínua, consumindo serviços de dados e dados compartilhados.

Saiba mais sobre o Snowflake para data lakes [aqui](#).



SOBRE O SNOWFLAKE

O Snowflake permite que todas as empresas impulsionem seus dados, graças ao Snowflake Data Cloud.

Os clientes usam o Data Cloud para eliminar silos de dados, descobrir e compartilhar dados com segurança, capacitar aplicativos de dados e executar inúmeras cargas de trabalho analíticas e de IA/ML. Onde quer que os dados ou os usuários estejam, o Snowflake proporciona uma única experiência de dados em inúmeras nuvens e regiões. Milhares de clientes em diversos setores, incluindo 639 das empresas que aparecem na Forbes Global 2000 (G2K) de 2023 (dados de 31 de julho de 2023), usam o Snowflake Data Cloud para impulsionar seus negócios. Saiba mais em [snowflake.com](https://www.snowflake.com).

