



SNOWFLAKE PARA CIÊNCIA DE DADOS

INTRODUÇÃO: TUDO GIRA EM TORNO DOS DADOS

As tecnologias de aprendizado de máquina (machine learning, ML) chegaram ao mercado. De acordo com uma pesquisa de 2019 da TDWI sobre o uso de inteligência artificial (IA) e ML, 92% dos participantes relataram usar tecnologia de aprendizado de máquina, e 85% afirmaram estar desenvolvendo modelos preditivos usando ferramentas para ML.¹

Os cientistas de dados precisam de grandes volumes de dados para desenvolver e treinar esses modelos de aprendizado de máquina. Na era da IA, ter acesso rápido e preciso aos dados tornou-se um diferencial competitivo fundamental. O gerenciamento de dados (descoberta, acesso seguro, limpeza, combinação e preparação para análise de dados) é geralmente visto como o aspecto mais demorado do processo.

Uma plataforma de dados eficiente é essencial

De acordo com a Forbes, os cientistas passam até 80% do tempo procurando, obtendo, consolidando, limpando e preparando dados para análises e treinamento.² O mesmo estudo revelou que cientistas de dados não só passam a maior parte do tempo gerenciando dados em vez de minerá-los ou modelá-los, como 76% desses profissionais altamente capacitados consideram a fase de preparação como a menos interessante de seu trabalho.³

Este white paper ajudará você a identificar os requisitos de dados que conduzem as iniciativas atuais de ciência de dados e ML. Ele também explicará como você pode cumprir tais requisitos fazendo uso de uma plataforma compatível com ferramentas líderes do setor da Snowflake e seus parceiros.

A PLATAFORMA MAIS COMPLETA PARA CIÊNCIA DE DADOS

A plataforma Snowflake combina o poder do data warehouse, a flexibilidade das plataformas de grandes volumes de dados, a elasticidade da nuvem e o compartilhamento de dados ativos a uma fração do custo das soluções tradicionais. O Snowflake oferece a simultaneidade, a simplicidade e o desempenho necessários para armazenar e analisar todos os seus dados em um só local, tanto para uso interno como para criar uma troca de dados. Milhares de clientes estão realizando a padronização nessa plataforma, pois ela atende a três necessidades essenciais:

- **Uma única fonte consolidada para todos os dados:** o Snowflake ajuda cientistas de dados a acessar dados estruturados e semiestruturados de uma fonte consistente, tornando fácil encontrar, consolidar, limpar e usar mais dos ativos de dados da sua organização. Os resultados de ciência de dados podem ser incorporados facilmente ao Snowflake para que os usuários empresariais possam acessá-los.
- **Preparação de dados eficiente e ágil:** o Snowflake oferece armazenamento virtual eficiente e dedicado capaz de ingerir, transformar e consultar dados por meio de SQL sem afetar outros usuários ou departamentos. Em diversos casos, o SQL no Snowflake é 10 vezes mais eficiente ao preparar dados em relação a ferramentas como o Spark, reduzindo a latência entre tarefas de ML.
- **Um amplo ecossistema de parceiros:** o Snowflake possui conectores para todas as tecnologias de ciências de dados estabelecidas e emergentes. Isso permite que os clientes escolham as melhores ferramentas de acordo com suas necessidades, e estas, por sua vez, acessam uma plataforma de dados unificada e consistente. O Snowflake exporta dados com facilidade para o Amazon S3 e outros armazenamentos de blob para garantir acesso universal às ferramentas de ciências de dados.

Cargas de trabalho Snowflake



Figura 1: o Snowflake possibilita muitos casos de uso e cargas de trabalho além da ciência de dados, então sua empresa pode aproveitar o poder de uma única plataforma para dados, análise de dados e análise preditiva.

PRINCIPAIS CONCEITOS E PERFIS DE CIÊNCIA DE DADOS

Os cientistas de dados usam tecnologia de aprendizado de máquina para identificar padrões, relações, correlações, resultados e inferências em seus dados. Essas descobertas baseadas em dados são incorporadas a modelos capazes de detectar fraude, prever ciclos de manutenção, mitigar perda de clientes, prever vendas e automatizar muitas outras tarefas voltadas para o futuro. As principais funções e perfis desse processo incluem:

- **Os cientistas de dados**, que desenvolvem modelos e os treinam com dados. Eles usam notebooks como Jupyter e Zeppelin, e linguagens como R, Python, Java e Scala.
- **Os analistas/cientistas de dados**, que usam esses modelos para realizar análise preditiva e prescritiva voltadas para a tomada de decisões empresariais, com base em seu conhecimento prático de aprendizado de máquina.
- **Os engenheiros de dados**, que preparam dados e estabelecem pipelines automatizados que alimentam os modelos de ML de forma contínua.

A FUNÇÃO DO APRENDIZADO DE MÁQUINA NA CIÊNCIA DE DADOS

O aprendizado de máquina lida principalmente com o aspecto de modelagem de dados na disciplina de ciência de dados, que é muito mais ampla, e engloba a preparação, a descoberta, a análise e a modelagem de dados. As ferramentas atuais de ML e ciência de dados conseguem lidar com muitos aspectos de análise, gerando modelos preditivos e prescritivos, colocando os modelos em produção e mantendo-os ao longo do tempo. As aplicações de análise preditiva e prescritiva muitas vezes são capazes de tomar decisões próprias sem intervenção humana, como é o caso do monitoramento de padrões de navegação na web para recomendar produtos e serviços a visitantes.

Os cientistas de dados usam ferramentas de análise de dados para formular uma hipótese e, em seguida, usam as linguagens de programação e bibliotecas de ML para criar suas previsões. Entre os tipos de ML estão regressão linear, regressão logística, classificação, árvores de decisão, aprendizado profundo e muitos outros. Algumas bibliotecas populares de ML incluem XGBoost, TensorFlow, scikit-learn e PyTorch.

Enquanto os cientistas de dados são responsáveis por criar e treinar modelos capazes de previsões confiáveis, os engenheiros cuidam de pipelines que alimentam modelos de ML com dados necessários para inferência. Os resultados desses processos de ML/IA são disponibilizados aos usuários empresariais para tomada de decisão baseada em dados.

O PROCESSO DO APRENDIZADO DE MÁQUINA

Iniciativas bem-sucedidas de aprendizado de máquina dependem que os modelos certos recebam os dados adequados no momento oportuno. Isso nem sempre é fácil, já que a maioria dos ciclos de aprendizado de máquina consiste em várias etapas, desde a descoberta e o desenvolvimento até a produção. Os dados são adicionados e preparados várias vezes em cada estágio do ciclo de ML, geralmente com diferentes requisitos de dados. O sucesso de ML parte do pressuposto que os dados certos, nas condições adequadas, serão fornecidos às plataformas de análises de dados apropriadas a fim de gerar resultados de negócios.

Como ilustrado na Figura 2, os cientistas de dados iniciam o processo ao encontrar, coletar, compreender e preparar os dados (etapas 1 a 3). É possível usar ferramentas de inteligência de mercado para entender melhor os dados e formular as hipóteses. Os cientistas testam muitos conjuntos de dados ao longo desse processo iterativo. Quando eles ampliam o escopo do conjunto, é preciso esperar que os engenheiros carreguem e preparem os dados. Isso causa atrasos e adiciona latência significativa entre cada iteração. Eles também precisam “moldar” os dados para normalizá-los, já que muitos algoritmos requerem formatos diferenciados.

A seguir, os dados de treinamento são executados nos modelos preparados nas etapas anteriores (4), avaliando os resultados para determinar a eficácia de cada modelo. Em seguida, os modelos são ajustados através de ciclos de engenharia de funcionalidades (3) e ajustes de hiperparâmetros (4).

Os modelos treinados resultantes são implementados na produção (5) para fornecer aos usuários empresariais ferramentas preditivas e prescritivas. Após a implementação na produção, os modelos são avaliados continuamente (6) para identificar desvios e determinar se estão obsoletos. É necessário retreinar os modelos periodicamente com novos dados. Essas atualizações de modelo representam outra iteração do ciclo de ML, exigindo o processamento de mais dados, o que demanda tempo e está sujeito a erros. Dependendo do caso de uso, pode ser necessário retreinar modelos com frequência de horas, dias ou semanas.

Fluxo de trabalho de ML

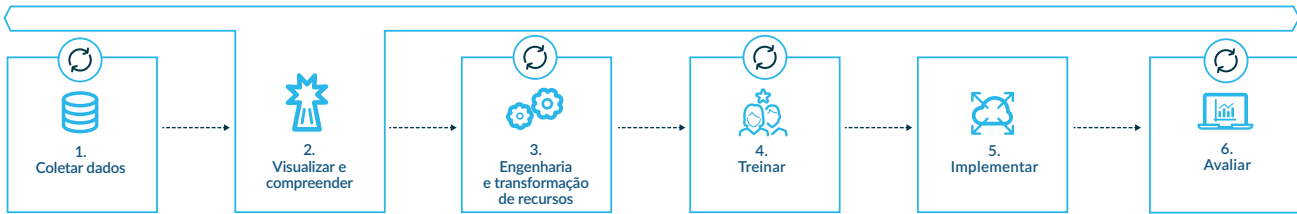


Figura 2: os dados conduzem o processo de ML, desde a coleta e a preparação até o treinamento, as previsões e a produção.

A FUNÇÃO DO DATA CLOUD

O aprendizado de máquina é uma atividade com uso intenso de dados, e o sucesso de cada modelo preditivo depende de grandes volumes de dados variados a serem coletados, preservados, transformados e apresentados de diversas formas. Isso envolve grandes volumes de dados caracterizados por muitas dimensões e detalhes, vindos de diversos contextos. Por exemplo, se você criou um modelo de aprendizado de máquina para prever a perda de clientes, é provável que tenha dados históricos e em tempo real sobre o comportamento dos clientes em relação a vendas, serviço, compra e interações com aplicações.

O Snowflake Data Cloud permite consolidar dados de data warehouses, data marts e data lakes em uma única fonte de verdade que conduz diversos tipos de análises e aplicações de ciência de dados. Isso facilita o compartilhamento de dados governados entre diversas equipes, interna ou externamente, ao possibilitar que os integrantes das equipes colaborem sem a necessidade de copiar ou mover dados de um lugar para outro. Os fluxos de trabalho de ciência de dados podem facilmente descobrir dados brutos, estruturados e semiestruturados, e acessá-los imediatamente graças ao suporte nativo a JSON, AVRO, XML, ORC e Parquet.

Poder usar um único conjunto de ferramentas para gerenciar dados estruturados e semiestruturados encurta o ciclo de descoberta e preparação de dados. Além disso, os dados gerados a partir dos algoritmos de ML são devolvidos ao repositório para que os usuários empresariais possam acessá-los, junto com os dados de origem. Ou seja, os dados estão sempre atualizados e mantidos de forma consistente para usuários empresariais, analistas e cientistas de dados.

Benefícios do Data Cloud

A TDWI recomenda adquirir uma plataforma de dados moderna, criada para a nuvem, que possa atender a todo o ciclo de aprendizado de máquina, inteligência artificial e desenvolvimento de aplicações preditivas. O que buscar em uma plataforma desse tipo? Para preparação de dados, é necessário poder trabalhar com grandes conjuntos de dados e tempos de resposta interativos. Para treinamento, é necessário percorrer os conjuntos de dados de modo iterativo. Para produção, é necessário contar com um pipeline de dados confiável, repetível e escalável.

Listamos a seguir alguns dos principais motivos para usar o Snowflake em seus projetos de ciência de dados:

- **Simplicidade:** não é necessário gerenciar várias plataformas computacionais nem fazer manutenção constante das integrações.
- **Segurança:** uma cópia dos dados é armazenada com segurança no ambiente do Snowflake, as credenciais de usuário são cuidadosamente gerenciadas e todas as transmissões são criptografadas.
- **Desempenho:** os resultados da consulta são armazenados em cache e podem ser repetidos durante o processo de ML, e nas análises de dados.
- **Isolamento de cargas de trabalho:** cada usuário e carga de trabalho pode receber recursos computacionais dedicados.
- **Elasticidade:** em segundos, é possível ampliar a capacidade para acomodar grandes tarefas de processamento de dados e reduzir facilmente assim que elas são concluídas, reduzindo custos com o pagamento por segundo de uso.
- **Suporte a dados estruturados e semiestruturados:** carregue, integre e analise todos os tipos de dados com facilidade dentro de um repositório unificado.
- **Simultaneidade:** execute cargas de trabalho simultâneas de grande porte em escala com diferentes dados compartilhados.

CONSOLIDAÇÃO DE DADOS PARA APRENDIZADO DE MÁQUINA E ANÁLISE

Existem diversas formas de provisionar dados para aplicações de aprendizado de máquina, sendo que a flexibilidade é essencial. Por exemplo, algumas organizações usam um data lake em conjunto com um data warehouse. Isso permite armazenar grandes volumes de dados brutos em formato nativo, a ser trabalhado para diversos tipos de análise quando necessário. A maioria das ferramentas de ciência de dados usa data lakes como fonte de dados, mas as estratégias de análise atuais cada vez mais empregam arquiteturas de dados multiplataforma que combinam grandes plataformas de dados, nuvens, data lakes e data warehouses. Muitas organizações líderes optam por deixar o data lake de lado e, em vez disso, consolidar seus dados integralmente em uma plataforma de dados na nuvem. Essa abordagem elimina a complexidade de gerenciar um data lake separado e também a necessidade de ter um pipeline de transformação de dados entre o data lake e o data warehouse. Com um repositório unificado, baseado em uma plataforma de dados na nuvem versátil, é possível escolher o armazenamento, o processamento e a economia para cada conjunto de dados e carga de trabalho, otimizando as opções de ML e análise de dados.

Assim que seus dados tiverem sido coletados e preparados, é preciso descobrir padrões e insights através de ferramentas de análise e análise de dados preditiva. O Snowflake permite combinar análise geral com análise preditiva, de forma que suas ferramentas de inteligência de mercado e de ciência de dados tenham uma visão consistente dos mesmos dados governados. Todas as ferramentas de ciência de dados usam as mesmas definições de dados como referência, permitindo que você reproduza consistentemente o conteúdo de consultas, previsões, painéis e relatórios. Tanto os dados brutos quanto os resultados de ML ficam na plataforma de dados para facilitar o acesso. Essa abordagem unificada permite que cientistas de dados retornem os resultados de atividades de aprendizado de máquina à plataforma de dados para análise geral, além de incluí-los no processo de tomada de decisão.

Com semântica, definições e modelos de dados em comum, todos estão em sintonia. Por exemplo, um gerente de vendas pode consultar um relatório de inteligência de mercado (business intelligence, BI) que apresenta o desempenho histórico da equipe de vendas. Já um modelo de ML poderia prever resultados esperados de vendas para os próximos trimestres com base na tendência das contas-alvo, e destacar a receita agendada e prevista no mesmo relatório.

ENGENHARIA DE DADOS AUTOMATIZADA, INTEGRAÇÃO DE DADOS E MODELAGEM DE DADOS

Ter sucesso com ML significa criar pipelines de dados eficientes e confiáveis que oferecem dados precisos e oportunos aos usuários empresariais, e também popular as aplicações e serviços usados por eles.

Ingestão de dados

O Snowflake inclui um serviço de ingestão sem servidor chamado Snowpipe que carrega dados de modo assíncrono e os disponibiliza imediatamente. As tarefas manuais de nivelamento de dados são totalmente automatizadas: a plataforma transforma os dados para o tipo e o formato exigidos por cada tabela de destino.

Conectores e adaptadores padrão permitem ingerir facilmente fluxos de eventos do Kafka e outros sistemas de mensagens, já que os fluxos e as tarefas do Snowflake facilitam o agendamento de carregamentos de dados para trabalhos SQL.

Ao “produtizar” o modelo de ML com um serviço de ingestão de dados automatizado, o pipeline simplifica as complexas tarefas de integração de dados. Os cientistas de dados podem encontrar e preparar dados conforme a demanda, sem esperar dias ou horas entre testes. Assim que um serviço de pipeline de dados automatizado entra em produção, os dados brutos são disponibilizados imediatamente sem precisar de ETL de um data lake. Conforme os dados são recebidos, eles passam pelo modelo automaticamente para gerar previsões. E como tudo é baseado na nuvem, os cientistas de dados podem usar recursos computacionais de virtual warehouse dedicados sem afetar outros usuários.

Recursos SQL universais

Os clientes Snowflake podem aproveitar uma fonte central de verdade com recursos SQL universais que potencializam as cargas de trabalho ETL e ELT robustas e eficientes. As transformações de dados com SQL são mais rápidas, simples e menos dispendiosas do que as mesmas operações realizadas com Spark. Já que os dados podem ser transformados como parte de uma consulta SQL, a transformação vira parte da análise. Tendo em vista a arquitetura e a compactação do Snowflake, é possível ingerir grandes volumes de dados em streaming rapidamente, que serão armazenados indefinidamente por um custo nominal. Diversas ferramentas de integração estão disponíveis para os engenheiros de dados, como Alteryx, Alooka, Matillion, Fivetran, Alation, Informatica e muitas outras.

Recursos de computação dedicados

Com o Snowflake, as cargas de trabalho de preparação de dados, gerenciamento de dados e ingestão de dados de ML contam com recursos dedicados que não afetam as cargas de trabalho de análise de dados e engenharia de dados não relacionadas a ML. Ao eliminar a contenção de recursos, é possível ingerir dados ativos e transformá-los em fluxo e disponibilizá-los imediatamente para análise. É possível personalizar o tamanho do seu data warehouse para cada carga de trabalho, ampliar conforme necessário e desativar os serviços de nuvem após a conclusão. Graças à escala linear, é possível solicitar o número exato de recursos necessários para executar consultas em um prazo previsível. Com elasticidade instantânea e cobrança por segundo, cada usuário e grupo de trabalho paga apenas pelos recursos de computação que usam. Em suma, essa arquitetura permite maximizar o desempenho e a eficiência de cada equipe enquanto fornece dados consistentes.

Segurança de dados robusta

Muitos projetos de ciência de dados legados dependem de Apache Hadoop, uma estrutura de código aberto para armazenagem e processamento de dados em uma estrutura distribuída. A arquitetura Hadoop, porém, oferece apenas controles de acesso rudimentares, não tendo sido projetada para atender às normas do setor relacionadas a segurança e privacidade de dados, incluindo HIPAA, PCI DSS e GDPR. Outros projetos de ciência de dados empregam armazenamentos de objetos gerais, como o Amazon S3, que também pecam na segurança.

Já o Snowflake Data Cloud é baseado em segurança multicamadas com criptografia, controle de acesso, monitoramento de rede e medidas de segurança física, além de monitoramento abrangente, alertas e práticas de segurança cibernética. Além de certificações de tecnologia padrão do setor como ISO/IEC 27001 e SOC 1/SOC 2 Type 2, o Snowflake atende à regulamentação relevante do setor e governamental, como PCI DSS, HIPAA/Health Information Trust Alliance (HITRUST) e FedRAMP. Os clientes Snowflake podem acessar, com segurança, dados para todo tipo de atividade em ciência de dados.

COMPARTILHAMENTO DE DADOS EFICIENTE

A plataforma Snowflake simplifica a troca de dados entre parceiros, fornecedores, vendedores e clientes por meio do Snowflake Data Marketplace e Data Exchange. Isso proporciona acesso a conjuntos de dados exclusivos que podem melhorar a eficácia dos modelos e oferecer possibilidades adicionais de engenharia de recursos. O compartilhamento seguro de dados no Snowflake não requer transferência de dados por FTP ou configuração de APIs para conectar aplicações. O processo simplifica a integração de ETL e sincroniza automaticamente dados "ativos" entre provedores e consumidores de dados. Como os dados de origem são compartilhados, em vez de copiados, os consumidores não precisam de armazenamento de nuvem extra. O Snowflake Data Marketplace e Data Exchange permitem que cientistas de dados colaborem em modelos com facilidade ao compartilhar dados brutos e processados.

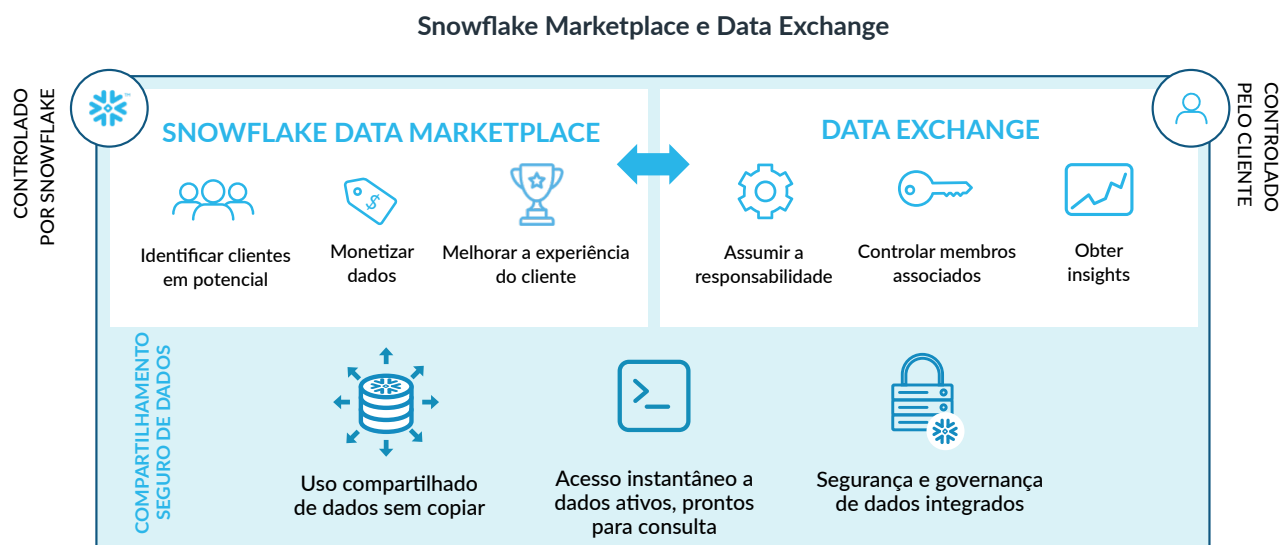


Figura 3: o Snowflake Secure Data Sharing possibilita o compartilhamento de dados externamente através do Snowflake Data Marketplace e a criação do seu próprio Data Exchange com clientes, fornecedores e demais parceiros de negócios.

AMPLO ECOSISTEMA DE PARCEIROS

O setor de aprendizado de máquina está evoluindo rapidamente, com novas ferramentas surgindo a cada ano. Através do amplo ecossistema de parceiros da Snowflake, os clientes podem aproveitar conexões diretas com todas as ferramentas, plataformas e linguagens de ciência de dados existentes e emergentes, como Python, R, Java e Scala; bibliotecas de código aberto, como PyTorch, XGBoost, TensorFlow, scikit-learn; notebooks, como Jupyter e Zeppelin, além de plataformas como Data Robot, Dataiku, H2O.ai, Amazon SageMaker, dentre outras. Ao fornecer um repositório de dados único e consistente, o Snowflake elimina a necessidade de retrabalhar os dados subjacentes sempre que você alternar ferramentas, linguagens ou bibliotecas. Além disso, o resultado dessas atividades é facilmente alimentado no Snowflake, disponibilizado por usuários não técnicos para gerar valor de negócios.

Ferramentas de ML baseadas em notebooks

Notebooks tradicionais de ML, como Jupyter e Zeppelin, potencializam as principais ferramentas de ciência de dados atuais, incluindo Amazon SageMaker, Dataiku, Zepl e muitas outras. Com isso, os cientistas de dados têm o máximo de controle sobre as estruturas e algoritmos escolhidos, podem conduzir engenharia aprofundada de funcionalidades, ajustar hiperparâmetros e criar, avaliar e produtizar modelos de ML de forma iterativa. Assim, é possível transformar intuições em previsões precisas ao testar iterações de algoritmos, pontuar o desempenho, escolher e refinar novos modelos. Usuários do Amazon SageMaker podem usar o Snowflake Python Connector para popular diretamente Pandas DataFrames. Essa conexão de alta velocidade leva a um treinamento mais ágil, assim como uma preparação de dados e ciclo de engenharia de funcionalidades otimizados que usam todo o potencial do ANSI SQL.

Ferramentas de AutoML

Como alternativa, ferramentas de AutoML como RapidMiner, BigSquid, H2o.ai e DataRobot podem automaticamente selecionar algoritmos, conduzir treinamento de modelos e escolher o melhor modelo. Essas ferramentas são uma ótima maneira de democratizar o acesso a análise de dados avançada, permitindo que analistas desempenhem funções

de ML sem grandes habilidades em programação ou conhecimento aprofundado de matemática e estatística. Algumas outras ferramentas aproximam essas duas abordagens, possibilitando que cientistas de dados personalizem processos de AutoML. DataRobot, líder no segmento de AutoML, possui integração nativa com o Snowflake, e seus usuários podem conectar rapidamente a conta do DataRobot ao Snowflake e usá-lo como armazenamento de dados.

Parceiros de nuvem e análise de dados

Independentemente do método de ML de sua preferência, o Snowflake permite o consumo de resultados através de painéis, relatórios e ferramentas de análise de negócios, aproveitando as conexões com outros parceiros de ecossistema, como Tableau, Looker, ThoughtSpot e Sigma. Além disso, o Snowflake permite armazenar e replicar seus dados em qualquer região e nuvem, incluindo em produtos populares da Amazon, Microsoft e Google. O Snowflake é capaz de exportar dados continuamente para tabelas externas em Amazon S3, Azure Blob e Google Cloud Storage, promovendo acesso universal de qualquer ferramenta. Por exemplo, é possível usar o Snowflake para complementar seu data lake no AWS, e conectar ao Amazon SageMaker para desenvolver, testar e implementar modelos de ML em escala. A plataforma automatiza tudo, do armazenamento de dados e processamento ao gerenciamento de transações, segurança, governança e gerenciamento de metadados.

Comece em minutos

Caso esteja procurando pela forma mais rápida de começar a usar o Snowflake e ML, considere o programa Snowflake Partner Connect, que simplifica a implementação por meio de integrações pré-configuradas com parceiros de tecnologia selecionados. É possível provisionar e configurar automaticamente aplicações de parceiros em minutos e carregar dados no Snowflake para uso imediato.

UM BOM EXEMPLO

A ConsumerTrack é uma editora e agência de publicidade digital que agrega e distribui dados de desempenho de sites de centenas de provedores para portais como CNN e MSN. Antes, sua equipe de ciência de dados lidava com um ambiente de ML que usava MySQL e diversas ferramentas de orquestração, levando a gargalos e problemas de latência com dados.

A ConsumerTrack ampliou seu data lake com o Snowflake e escolheu o Amazon SageMaker como serviço inteiramente gerenciado para automatizar o fluxo de trabalho de ML. Ele rotula e prepara dados, escolhe um algoritmo, treina um modelo, ajusta e otimiza o modelo para implementação, faz previsões e por fim entra em ação.

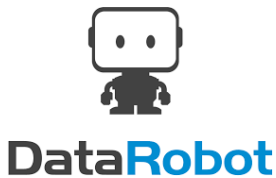
Agora os dados fluem para o data lake através de um pipeline automatizado que usa AWS Lambda e AWS Glue. Os dados são selecionados e carregados no Snowflake, e os fluxos de dados são configurados com alertas personalizados. O Amazon SageMaker conecta-se ao Snowflake para simplificar o desenvolvimento, a testagem e a criação de modelos de ML.

Assim, a ConsumerTrack eliminou gargalos, e o tempo de obtenção de informações caiu de horas para minutos. O Snowflake reduz substancialmente o tempo gasto em descoberta e preparação de dados. Seu ecossistema amplo permite que a ConsumerTrack se conecte a diversos tipos de plataformas e ferramentas de ciência de dados, incluindo um conector nativo para Python. Quando necessário, a equipe de ciência de dados pode exportar os dados para qualquer armazenamento blob para acesso universal.

PRÓXIMO PASSO

Para obter mais informações sobre aprendizado de máquina, visite [a página de ciência de dados de Snowflake](#) e [a página da plataforma Snowflake](#).

Parceiros de ciência de dados em destaque



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 590 das empresas que aparecem na Forbes Global 2000 (G2K) de 2022 (dados de 30 de abril de 2023), usam o Snowflake Data Cloud para impulsionar seus negócios. Saiba mais em snowflake.com.



© 2021 Snowflake Inc. Todos os direitos reservados. Snowflake, o logotipo da Snowflake e todos os demais nomes de produtos, recursos e serviços da Snowflake mencionados neste documento são marcas registradas ou marcas comerciais da Snowflake Inc. nos Estados Unidos e em outros países. Todos os outros nomes de marcas ou logotipos mencionados ou usados neste documento são apenas para fins de identificação e podem ser marcas comerciais de seus respectivos detentores. A Snowflake não pode ser associada a tais detentores, nem patrocinada ou apoiada por eles.

CITAÇÕES

¹ "Best Practices Report: Driving Digital Transformation Using AI and Machine Learning" (tdwi.org/bpreports).

² Forbes "Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says" (bit.ly/38EbXmN).

³ Forbes "Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says" (bit.ly/38EbXmN).