

Introdução

No cenário competitivo atual, a utilização estratégica de dados tornou-se essencial para o crescimento sustentável e a diferenciação no mercado. A **Adventure Works (AW)**, uma indústria de bicicletas em franco crescimento, reconhece a importância de transformar-se em uma empresa orientada por dados (data-driven) para manter seu ritmo de expansão e superar a concorrência. Com um portfólio robusto de mais de 500 produtos distintos, atendendo a 20.000 clientes e gerenciando 31.000 pedidos, a AW busca alavancar suas informações para nortear decisões estratégicas e operacionais.

A diretoria da AW identificou uma série de questões que precisam ser respondidas por meio do cruzamento de dados, orientando assim o desenvolvimento das tabelas de fatos e dimensões no data warehouse. Inicialmente, o projeto focou na área de vendas (sales) para obter resultados rápidos, embora reconheça a necessidade de integrar dados de outras áreas para uma visão holística do negócio. Durante o diagnóstico inicial, foram identificados sistemas internos que geram dados relevantes e que deverão ser incorporados à infraestrutura de dados da empresa.

Entretanto, o projeto enfrenta desafios internos, incluindo a falta de consenso sobre o cronograma, os custos envolvidos e a percepção de retorno sobre o investimento. Enquanto o diretor de inovação, João Muller, e o CEO, Carlos Silveira, apoiam vigorosamente a iniciativa, a diretora comercial, Silvana Teixeira, questiona a alocação de recursos para o projeto, preferindo investimentos em ações promocionais de impacto imediato. Além disso, há uma desconfiança quanto à eficácia das soluções data-driven propostas anteriormente por fornecedores de sistemas como CRM e Web Analytics.

Para garantir o acesso eficiente aos dados, o diretor de TI, Nilson Ramos, designou o analista Gabriel Santos, cuja disponibilidade limitada exige uma coordenação ágil e eficiente das interações para que todos os dados necessários estejam prontamente disponíveis. Adicionalmente, João Muller identificou um potencial patrocinador em Luís Soares, gestor da equipe de Planejamento de Demanda, que busca aprimorar o processo atual baseado em médias móveis de 3 meses utilizando Excel, demonstrando interesse em soluções preditivas que justifiquem um investimento em inovação.

Diante desse contexto, a solução foi desenvolvida através das seguintes etapas:

1. **Diagrama Conceitual do Data Warehouse (DW):** Criação de uma representação visual das principais entidades e relacionamentos para estruturar a base de dados.
2. **Implementação do Data Warehouse na Nuvem (Snowflake):** Utilização da plataforma Snowflake para garantir escalabilidade, segurança e acessibilidade dos dados.
3. **Transformação dos Dados com dbt Cloud:** Aplicação de ferramentas de transformação de dados para garantir a qualidade e a integridade das informações.

4. **Versionamento no GitHub:** Gerenciamento de versões e colaboração eficiente através do uso do GitHub.
5. **Desenvolvimento de Dashboards:** Criação de visualizações estratégicas para o CEO, com indicadores chave, e relatórios operacionais focados em pedidos, regiões e clientes.
6. **Desenvolvimento de Modelos de Machine Learning:** Implementação de previsões de demanda utilizando séries temporais, hierarquias e modelos de regressão, visando aprimorar o planejamento de produção.

Este relatório documenta detalhadamente cada uma dessas etapas, demonstrando como a Adventure Works está se transformando em uma organização data-driven, superando desafios internos e aproveitando suas vastas bases de dados para impulsionar decisões estratégicas e operacionais informadas.

Desenvolvimento do Data Warehouse

O desenvolvimento do Data Warehouse (DW) para a Adventure Works foi estruturado de maneira meticulosa, utilizando práticas modernas de engenharia de dados como a Modern Data Stack e ferramentas robustas para garantir escalabilidade, integridade e acessibilidade das informações. A seguir, detalhamos cada componente e etapa envolvida na construção do DW, conforme implementado no ambiente dbt Cloud.

Estrutura de Arquivos no dbt Cloud

O projeto do Data Warehouse foi organizado no dbt Cloud seguindo uma estrutura de diretórios clara e modular, facilitando a manutenção, colaboração e escalabilidade do projeto. A estrutura de arquivos está organizada da seguinte forma:

Componentes Principais

1. Configuração do Projeto (dbt_project.yml)

O arquivo **dbt_project.yml** contém as configurações fundamentais do projeto dbt, incluindo o nome do projeto, a versão do dbt, os diretórios de modelos, sementes, snapshots e testes, além de configurações específicas para cada um desses componentes. Este arquivo é essencial para definir como o dbt irá interpretar e executar os modelos de dados.

2. Macros

As macros são funções reutilizáveis escritas em SQL que permitem a criação de lógica parametrizada e abstrações dentro dos modelos. No projeto, a macro **generate_schema_name.sql** é utilizada para gerar nomes de esquemas dinâmicos, facilitando a organização e a segmentação dos dados no DW.

3. Modelos (models)

Os modelos são a espinha dorsal do DW, representando as tabelas e visualizações que compõem o Data Warehouse. A estrutura de modelos está dividida em duas principais categorias:

- **Staging (staging/erp):**

Esta camada é responsável por extrair e transformar dados brutos provenientes de sistemas ERP. Cada arquivo SQL na pasta **staging/erp** corresponde a uma tabela de origem, realizando a limpeza, padronização e preparação dos dados para serem carregados nas camadas superiores. Os arquivos **.yaml** associados definem as fontes de dados e documentam os modelos, facilitando o entendimento e a governança dos dados.

- **Marts (marts/comercial):**

Esta camada contém os modelos de dados específicos para a área comercial. Inclui tabelas de dimensões (**dim_**) e fatos (**fact_**) que são essenciais para análises de vendas. Exemplos de modelos implementados:

- **dim_customers.sql**: Tabela de dimensões de clientes, agregando informações detalhadas sobre os clientes da AW.
- **fact_sales.sql**: Tabela de fatos de vendas, registrando todas as transações realizadas.
- **agg_salesperson_salesregion.sql**: Modelo de agregação que relaciona vendedores com suas respectivas regiões de vendas.

- Cada modelo é acompanhado por arquivos **.yaml** que definem os esquemas, testes e documentações, assegurando a qualidade e a clareza dos dados.

4. Sementes (seeds)

As sementes são arquivos CSV que contém dados estáticos ou de referência que são carregados diretamente no DW. No projeto, a pasta **seeds/sap_adventure_works** está organizada por áreas de negócio, como **human_resources**, **person**, **production**, **purchasing** e **sales**. Exemplos de sementes:

- **department.csv**: Lista de departamentos da empresa.
- **product.csv**: Detalhes dos produtos disponíveis.
- **vendor.csv**: Informações sobre os fornecedores.

Estas sementes são utilizadas para preencher tabelas de dimensões e auxiliar nas transformações dos modelos.

5. Snapshots

Os snapshots permitem capturar o estado dos dados em pontos específicos no tempo, facilitando a análise de mudanças históricas. A pasta **snapshots** está preparada para armazenar scripts que definem como e quando capturar esses estados, garantindo a rastreabilidade das alterações nos dados ao longo do tempo.

6. Testes (tests)

Garantir a integridade e a qualidade dos dados é fundamental. A pasta **tests** contém scripts SQL que executam verificações específicas, como:

- **tst_total_valor_pago_produto.sql**: Verifica a consistência do valor total pago por produto.
- **tst_total_vendas.sql**: Assegura que o total de vendas está correto.
- Testes por ano (**tst_total_vendas_2011.sql**, etc.) para validar dados históricos.

Esses testes são executados regularmente para identificar e corrigir possíveis inconsistências ou erros nos dados.

7. Versionamento com GitHub

O projeto utiliza o GitHub para versionamento de código, facilitando a colaboração entre membros da equipe e garantindo que todas as alterações sejam rastreadas e auditáveis. Os arquivos **package-lock.yml** e **packages.yml** gerenciam as dependências do projeto, assegurando que todas as bibliotecas e pacotes necessários estejam disponíveis e atualizados.

Implementação na Nuvem com Snowflake

A escolha do Snowflake como plataforma de Data Warehouse na nuvem proporciona à Adventure Works uma solução escalável, segura e altamente disponível. O Snowflake facilita a integração com o dbt Cloud, permitindo a execução eficiente de transformações de dados e a gestão de recursos conforme a demanda. A arquitetura na nuvem também garante que os dados estejam acessíveis para análises em tempo real, suportando o crescimento contínuo da empresa.

Transformação de Dados com dbt Cloud

O dbt Cloud foi utilizado para orquestrar e automatizar as transformações de dados. Com o dbt, implementamos um fluxo de trabalho que inclui:

- **Modelagem de Dados**: Criação de modelos SQL para transformar dados brutos em informações estruturadas e significativas.
- **Testes de Qualidade**: Implementação de testes automatizados para validar a integridade e a consistência dos dados.
- **Documentação**: Geração de documentação interativa e atualizada dos modelos, facilitando o entendimento e a governança dos dados.

Versionamento e Colaboração com GitHub

O uso do GitHub para versionamento garante que todas as alterações no projeto sejam devidamente registradas e possam ser revertidas se necessário. A colaboração entre os membros da equipe é facilitada por meio de **branches**, **pull requests** e **revisões de código**, promovendo uma cultura de melhoria contínua e qualidade no desenvolvimento do DW.

Desenvolvimento de Dashboards

Para atender às necessidades de visualização e tomada de decisão da diretoria, foram desenvolvidos dashboards personalizados:

- **Dashboard Estratégico para o CEO:** Inclui indicadores chave de performance (KPIs) que oferecem uma visão geral da saúde e desempenho da empresa.
- **Relatórios Operacionais:** Focados em pedidos, regiões e clientes, fornecendo insights detalhados para a equipe operacional gerenciar eficientemente as vendas e identificar oportunidades de crescimento.

Esses dashboards foram implementados utilizando o PowerBI como ferramenta de visualização de dados que se conectam diretamente ao Data Warehouse, garantindo atualizações em tempo real e interatividade nas análises.

Modelos de Machine Learning para Previsão de Demanda

Uma das inovações estratégicas implementadas foi o desenvolvimento de modelos de Machine Learning para previsão de demanda. Utilizando técnicas de séries temporais, hierarquias e regressão, os modelos permitem:

- **Previsão de Demanda:** Antecipar a demanda futura com base em dados históricos e tendências.
- **Planejamento de Produção:** Otimizar a produção de bicicletas, evitando excessos ou faltas de estoque.
- **Análise de Cenários:** Avaliar diferentes cenários e seus impactos no planejamento estratégico.

Essas previsões auxiliam a equipe de Planejamento de Demanda a tomar decisões mais informadas e a alocar recursos de maneira eficiente, aumentando a competitividade da Adventure Works no mercado.

Integração com Ferramentas de Visualização

Os dashboards desenvolvidos para o CEO e para a equipe operacional foram integrados com ferramentas de visualização de dados com Power BI, que se conectam diretamente ao Snowflake. Essa integração permite a criação de visualizações interativas e dinâmicas, proporcionando uma experiência intuitiva para a análise de dados e a tomada de decisões estratégicas.