

Projeto de Dados: Adventure Works

Seu Nome

3 de fevereiro de 2025

Sumário

1	Introdução	2
2	Objetivos do Projeto	2
3	Contexto e Fontes de Dados	2
4	Arquitetura do Projeto	3
5	Pipeline de Dados	4
6	Mapeamento das Necessidades dos Stakeholders	5
7	Modelagem de Dados	5
8	Dashboards e KPIs	6
8.1	Primeiro Dashboard: Clientes e LTV	6
8.2	Segundo Dashboard: Vendas e Performance	7
8.3	Conclusões e Ações Possíveis	8
9	Conclusão	8

1 Introdução

A **Adventure Works (AW)** é uma indústria de bicicletas em crescimento que busca se transformar em uma empresa *data-driven*, utilizando dados para fundamentar suas decisões estratégicas. Atualmente, a empresa conta com:

- Mais de 500 produtos distintos;
- Aproximadamente 20.000 clientes;
- Cerca de 31.000 pedidos registrados.

O projeto tem como objetivo construir um pipeline de dados robusto, garantindo governança, escalabilidade e confiabilidade, e permitindo análises precisas e oportunas que apoiem as estratégias comerciais.

2 Objetivos do Projeto

Os principais objetivos são:

- Consolidar e organizar dados provenientes de diversas fontes (ERP, CRM, Web Analytics e Site);
- Criar uma infraestrutura escalável e confiável para análises;
- Gerar insights acionáveis sobre clientes, vendas e produtos;
- Automatizar a ingestão e transformação dos dados, reduzindo processos manuais e aumentando a eficiência.

3 Contexto e Fontes de Dados

O diagnóstico inicial identificou que os dados da AW provêm de quatro fontes principais:

- **ERP (SAP):** Gestão interna, produtos, fornecedores e pedidos de compra;
- **CRM (Salesforce):** Clientes, vendas e territórios comerciais;
- **Web Analytics (Google Analytics):** Análise de tráfego e comportamento do usuário;
- **Site (WordPress):** Dados de conteúdo digital e interações online.

Mapeamento das Tabelas e Fontes de Extração

Tabela	Função Associada	Fonte de Extração
salesorderdetail	CRM	Salesforce
salesorderheader	CRM	Salesforce
customer	CRM	Salesforce
person	CRM	Salesforce
store	CRM	Salesforce
productvendor	ERP	SAP
stateprovince	CRM / ERP	Salesforce / SAP
countryregion	CRM / ERP	Salesforce / SAP
salesterritory	CRM	Salesforce
purchaseorderdetail	ERP	SAP
purchaseorderheader	ERP	SAP
product	ERP	SAP
productcategory	ERP	SAP
productsubcategory	ERP	SAP
salesperson	CRM	Salesforce
employee	ERP	SAP
salesreason	Web Analytics	Google Analytics

4 Arquitetura do Projeto

O projeto foi desenvolvido com base no **Modern Data Stack (MDS)**, integrando as seguintes ferramentas e processos:

- **Extração e Carregamento:** Utilização do Fivetran para a ingestão automática dos dados, que são armazenados no Snowflake.
- **Transformação e Modelagem:** Uso do dbt (Data Build Tool) para transformar e modelar os dados, com testes automáticos, documentação integrada e organização dos modelos em camadas.
- **Orquestração:** Apache Airflow agenda e executa o pipeline. Uma DAG diária (às 17:00 UTC / 14:00 BRT) executa as tarefas de `dbt test` (para validar a integridade dos dados) e `dbt run` (para atualizar os modelos).

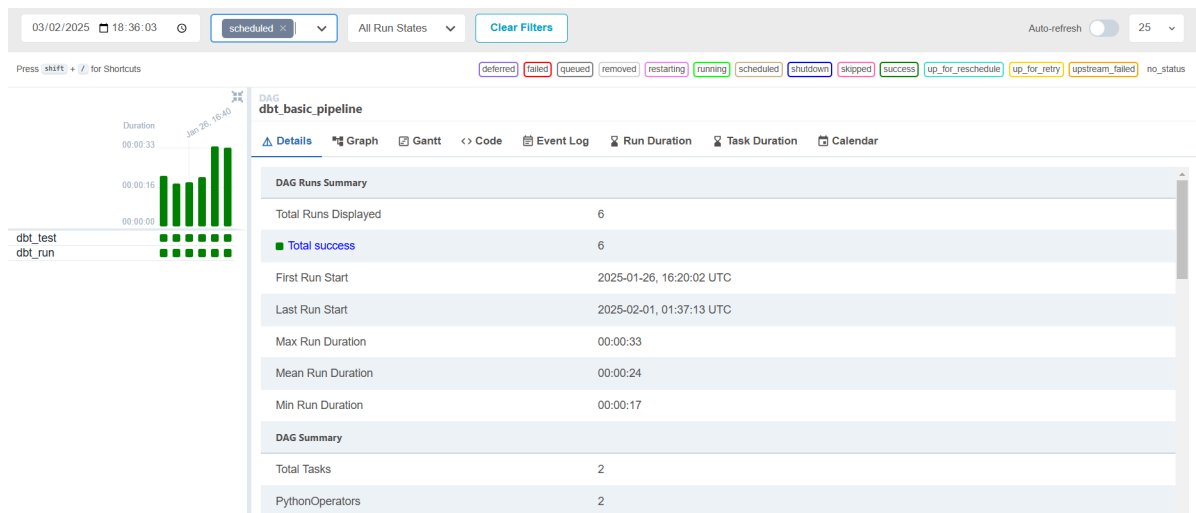


Figura 1: Execução da DAG 'dbt_basic_pipeline' no Apache Airflow. O resumo indica 6 execuções bem-sucedidas, com tempos de execução variando entre 17 e 33 segundos. Todas as tarefas 'dbt_test' e 'dbt_run' foram concluídas sem falhas.

Essa abordagem garante:

- **Automação:** Processos ETL/ELT automatizados e monitorados.
- **Escalabilidade:** Infraestrutura preparada para o crescimento do volume e complexidade dos dados.
- **Confiabilidade:** Processos documentados que asseguram a integridade e rastreabilidade dos dados.

5 Pipeline de Dados

O pipeline integra as etapas de extração, transformação e carregamento, garantindo a atualização contínua e a confiabilidade dos dados. Destacam-se:

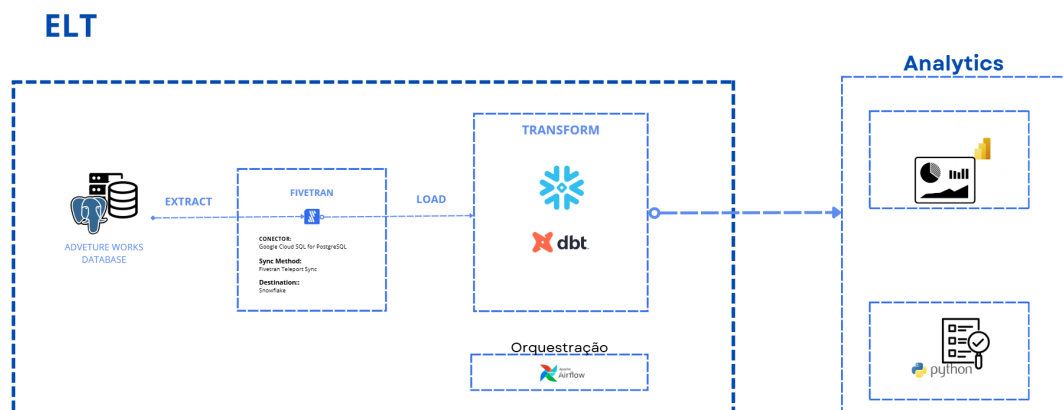


Figura 2: Diagrama do Pipeline de Dados

- **Extração e Carregamento:** Dados são extraídos via Fivetran e armazenados no Snowflake.

- **Transformação e Modelagem:** O dbt transforma e modela os dados com testes automáticos e documentação integrada.
- **Orquestração:** O Apache Airflow executa, diariamente, uma DAG que roda dbt test e, em seguida, dbt run.

6 Mapeamento das Necessidades dos Stakeholders

Para atender às diversas demandas, foi realizado um mapeamento detalhado das necessidades dos stakeholders. A tabela a seguir apresenta cada stakeholder, sua necessidade e a forma de atendimento proposta:

Stakeholder	Necessidade	Como Atender
CEO (Carlos Silveira)	KPIs estratégicos e insights rápidos.	Relatório gerencial com vendas totais, ticket médio e crescimento por região, utilizando ferramentas de BI.
Diretor de Inovação (João Muller)	Justificar investimentos futuros.	Relatório detalhado com métricas de desempenho e plano de expansão.
Diretora Comercial (Silvana Teixeira)	Demonstrar impacto comercial.	Apresentação de métricas operacionais (ticket médio, periodicidade de compras, métodos de pagamento) com recomendações estratégicas.
Diretor de TI (Nilson Ramos)	Infraestrutura escalável e acessível.	Implementação de pipelines ETL/ELT automatizados (ex.: com Airflow), com documentação e capacitação da equipe.
Analista de TI (Gabriel Santos)	Agilizar a extração e comunicação de dados.	Desenvolvimento de pipelines otimizados para evitar sobrecarga e melhorar a transferência dos dados.
Gestor de Planejamento (Luís Soares)	Comparar modelo preditivo com Excel.	Criação e validação de um modelo estatístico de previsão de demanda, comparando-o com o método atual.

7 Modelagem de Dados

Nesta seção, apresentamos a modelagem de dados realizada para a **Adventure Works**, que serve de base para a construção do Data Warehouse e suporta as análises apresen-

tadas nos dashboards.

O modelo conceitual (Figura 3) organiza as principais entidades e seus relacionamentos, permitindo a integração dos dados provenientes das diversas fontes.

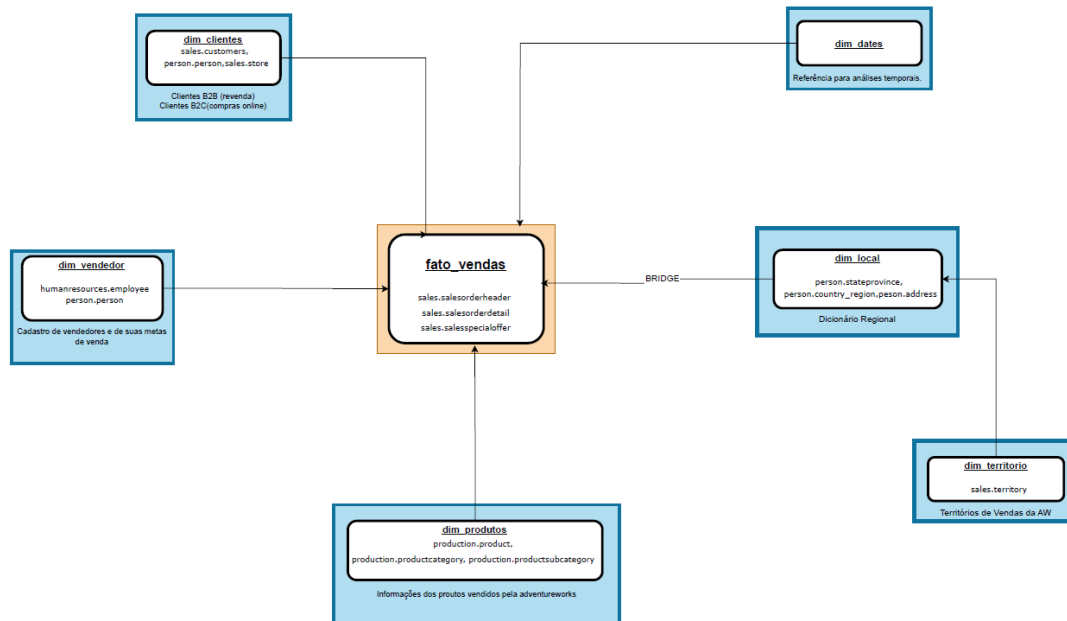


Figura 3: Modelo Conceitual do Data Warehouse

Essa abordagem modular, utilizando o dbt, garante que cada etapa do pipeline de transformação seja testada, documentada e organizada em camadas, facilitando a manutenção e a escalabilidade da solução.

8 Dashboards e KPIs

Os insights são apresentados em dashboards interativos no **Power BI**, que auxiliam os stakeholders na tomada de decisões estratégicas. Foram desenvolvidos dois dashboards principais:

8.1 Primeiro Dashboard: Clientes e LTV

Focado em clientes, LTV (Lifetime Value) e métricas de retenção, este dashboard exibe: **KPIs Principais (Cards no Topo):**

- **Clientes:** 19.119 – Total de clientes cadastrados.
- **Clientes Ativos:** 10.415 – Clientes que ainda realizam compras.
- **Receita Média por Cliente:** \$5,77 mil – Receita média gerada por cliente.
- **LTV:** \$12.681 – Valor total esperado que um cliente gera ao longo do tempo.
- **Tempo Médio de Vida:** 6 anos – Período médio que um cliente permanece ativo.

- **Taxa de Churn:** 45,53% – Percentual de clientes que deixaram de comprar.

Gráficos e Tabelas:

- **Volume Acumulado de Vendas por Dia da Semana:** Indica que as vendas são mais intensas no início da semana (segundas e terças), sugerindo um comportamento de compra predominante nesses dias.
- **Ticket Médio por Tipo de Cliente:** Revela que clientes do tipo B2C possuem um ticket médio de aproximadamente \$42 mil.
- **Clientes Ativos vs. Clientes Perdidos:** Gráfico de pizza que demonstra a proporção entre clientes ativos e os que deixaram de comprar.
- **Evolução da Receita Bruta e Média Móvel:** Exibe uma tendência de queda consistente na receita bruta ao longo dos anos.
- **LTV Acumulado por Tempo Desde a Aquisição:** Análise de coorte que mostra o crescimento do LTV com o tempo, identificando o ponto de estabilização.
- **Top Clientes – Revenda:** Lista os clientes com maior faturamento, útil para estratégias de fidelização.
- **Tempo Médio de Vida e Receita por Faixa Etária:** Evidencia que clientes mais jovens (<30 anos) possuem maior tempo de vida, com uma distribuição específica de receita.

8.2 Segundo Dashboard: Vendas e Performance

Este dashboard foca em vendas, performance dos vendedores e produtos mais vendidos, apresentando:

KPIs Principais (Cards no Topo):

- **Receita Bruta:** \$110,37 milhões – Valor total gerado pelas vendas.
- **Receita Líquida:** \$109,85 milhões – Receita após deduções.
- **Ticket Médio:** \$21,40 mil – Valor médio gasto por pedido.
- **Número de Pedidos:** 31.465 – Total de pedidos realizados.
- **Variação Anual da Receita Bruta:** -44,99% – Forte queda na receita ao longo do tempo.
- **Clientes:** 19.119 – Número total de clientes.

Gráficos e Tabelas:

- **Clientes por Território de Vendas:** Mapa geográfico que identifica regiões mais lucrativas.
- **Tendência de Vendas ao Longo do Tempo:** Exibe um pico de vendas por volta de 2012, seguido de declínio, com análise por categorias (Accessories, Bikes, Clothing, Components).

- **Performance de Vendas – Vendedores e Regiões:** Mapa de Sankey que mostra a contribuição de cada vendedor por território e categoria de produto.
- **Evolução do Ticket Médio e Receita YTD:** Gráfico que evidencia a evolução do ticket médio e da receita ao longo dos anos, com queda após o pico de 2012-2013.
- **Top 5 Produtos Mais Vendidos:** Lista dos produtos mais vendidos, como AWC Logo Cap, Water Bottle, Long-Sleeve Jersey e Helmet.
- **Top 5 Vendedores em Volume de Vendas:** Mostra a contribuição percentual dos principais vendedores, destacando os de melhor desempenho.

8.3 Conclusões e Ações Possíveis

A análise dos dashboards revela oportunidades e desafios:

- **Retenção de Clientes:** A alta taxa de churn (45,53%) indica a necessidade de reforçar estratégias de fidelização, como descontos e programas de recompra.
- **Queda na Receita:** A tendência de diminuição na receita sugere a necessidade de inovação em produtos e melhorias na estratégia de marketing.
- **Desempenho Regional e dos Vendedores:** A análise geográfica e por desempenho possibilita identificar regiões e vendedores mais lucrativos, orientando ajustes na força de vendas.
- **Oportunidade no Ticket Médio e LTV:** Embora o ticket médio seja elevado, estratégias de upsell e cross-sell podem prolongar o tempo de vida do cliente e aumentar o LTV.
- **Categorias de Produto em Declínio:** Analisar as categorias que apresentaram maior queda pode fornecer insights para ajustes estratégicos e ações corretivas.

9 Conclusão

O projeto de dados da **Adventure Works** estabeleceu uma base sólida para a tomada de decisões estratégicas, por meio de uma infraestrutura automatizada, escalável e confiável. A implementação do **Modern Data Stack** – utilizando Fivetran, Snowflake, dbt e Airflow – possibilitou a criação de um fluxo de dados eficiente, suportando análises avançadas e preditivas.

Além disso, os dashboards interativos desenvolvidos no Power BI oferecem insights detalhados sobre clientes, vendas e performance, permitindo ações estratégicas que podem impulsionar a retenção de clientes e otimizar as vendas.

Com essa estrutura, a empresa está bem posicionada para expandir suas análises, otimizar processos e aprimorar suas estratégias comerciais. Recomenda-se a continuidade do desenvolvimento, com foco na evolução da modelagem preditiva e na ampliação do escopo analítico, para maximizar o valor extraído dos dados.

⁰Este documento passou por revisões e foi construído com o auxílio de Inteligência Artificial (IA).