

Planejamento do Projeto de Dados

[ADVENTURE WORKS]

[NATACHA S. S. DA SILVA - ANALYTICS ENGINEER]

[03/2024]

Introdução

A Adventure Works é uma indústria de bicicletas em franco crescimento que se orgulha de possuir mais de 500 produtos distintos, 20.000 clientes e 31.000 pedidos. A empresa tem como visão se tornar uma referência no mercado de bicicletas, oferecendo produtos de qualidade, atendimento personalizado e entrega rápida.

Para alcançar essa visão, a empresa precisa utilizar seus dados de forma estratégica, norteando suas decisões com base em informações confiáveis, atualizadas e relevantes. Por isso, a empresa contratou um consultor especializado em projetos de dados para desenvolver e implantar uma solução de data warehouse, que permitirá a integração, a organização e a análise dos dados provenientes de diferentes fontes e sistemas.

O data warehouse é uma base de dados centralizada que armazena dados históricos e consolidados de diversas áreas da empresa, facilitando o acesso e a visualização dos dados por meio de ferramentas de business intelligence (BI). O data warehouse é composto por tabelas de fatos e dimensões, que seguem um modelo multidimensional de dados. As tabelas de fatos contêm as medidas ou indicadores de desempenho da empresa, como vendas, lucros, custos, etc. As tabelas de dimensões contêm os atributos ou características que descrevem as medidas, como produto, cliente, tempo, localização, etc.

O objetivo deste projeto é desenvolver e implantar um data warehouse para a área de vendas da Adventure Works, que é responsável por gerar a maior parte da receita da empresa. O data warehouse de vendas deverá responder a uma série de perguntas de negócio.

Para responder a essas perguntas, o data warehouse de vendas deverá integrar os dados provenientes de vários sistemas que a Adventure Works utiliza, tais como:

- Sistema de gestão empresarial (ERP), que contém os dados de cadastro de produtos, clientes, fornecedores, funcionários, etc.
- Sistema de gestão de relacionamento com o cliente (CRM), que contém os dados de interação com os clientes, como contatos, campanhas, oportunidades, etc.
- Sistema de análise da web (Web Analytics), que contém os dados de navegação e de conversão dos clientes no site da empresa.
- Sistema de rastreamento de pedidos (Order Tracking), que contém os dados de status e de localização dos pedidos em trânsito.

O projeto de data warehouse de vendas enfrenta alguns desafios, como:

- A falta de consenso entre as áreas de negócio sobre a necessidade e o valor do projeto, gerando resistência e desconfiança.
- A dificuldade de acesso e de qualidade dos dados, que exigem um trabalho de extração, transformação e carga (ETL) complexo e demorado.
- A limitação de recursos e de tempo, que demandam um planejamento e uma priorização eficientes das atividades e das entregas.

Objetivos esperados

O projeto de data warehouse de vendas da Adventure Works tem como principais objetivos:

- Criar uma fonte única e confiável de dados de vendas, integrando os dados de diferentes sistemas e garantindo a qualidade, a consistência e a segurança dos dados.
- Disponibilizar os dados de vendas de forma fácil e rápida para as áreas de negócio, por meio do Power BI, permitindo assim a criação de relatórios, dashboards e análises.
- Responder às perguntas de negócio, fornecendo insights e recomendações baseados em dados para apoiar a tomada de decisão.
- Aumentar a eficiência e a produtividade da área de vendas, reduzindo o tempo e o esforço gastos com a coleta, o tratamento e a manipulação dos dados.
- Melhorar o desempenho e a rentabilidade da área de vendas, identificando oportunidades de aumento de receita, redução de custos, otimização de processos e fidelização de clientes.

- Demonstrar o valor e o potencial do uso de dados para outras áreas da empresa, incentivando a cultura data driven e gerando demanda por novos projetos de dados.

Partes interessadas

As partes interessadas do projeto são:

- **Diretoria da empresa:** é composta pelo CEO, Carlos Silveira, pelo diretor de inovação, João Muller, pela diretora comercial, Silvana Teixeira, e pelo diretor de TI, Nilson Ramos. A diretoria é responsável por definir a visão, a missão e os objetivos estratégicos da empresa, bem como aprovar o orçamento e o escopo do projeto. A diretoria também é a principal beneficiária dos resultados e dos insights gerados pelo data warehouse de vendas, que devem apoiar a tomada de decisão e a melhoria contínua da empresa.
- **Consultor especializado em projetos de dados:** é o profissional contratado pela Adventure Works para desenvolver e implantar o data warehouse de vendas. O consultor é responsável por realizar o levantamento de requisitos, o planejamento, a modelagem, a construção, o teste, a implantação e a documentação do data warehouse, seguindo as melhores práticas e as metodologias de gerenciamento de projetos. O consultor também é responsável por treinar e orientar as áreas de negócio e de TI na utilização e na manutenção do data warehouse, bem como apresentar os benefícios e os resultados do projeto para a diretoria e as demais partes interessadas.
- **Analista de dados:** é o funcionário da área de TI da Adventure Works, Gabriel Santos, que foi indicado pelo diretor de TI, Nilson Ramos, para auxiliar o consultor no acesso aos dados. O analista de dados é responsável por administrar os bancos de dados e por fornecer os dados necessários para o projeto, seguindo as normas de segurança e de qualidade dos dados. O analista de dados também é responsável por validar e testar o data warehouse de vendas, bem como por apoiar o consultor na resolução de problemas técnicos e na implantação do data warehouse.
- **Equipe de vendas:** é composta pelos vendedores, gerentes e supervisores da área comercial da Adventure Works, que são responsáveis por realizar as vendas dos produtos da empresa para os clientes. A equipe de vendas é a principal usuária do data warehouse de vendas, que deve fornecer os dados e as informações necessárias para o planejamento, a execução e o controle das atividades de vendas. A equipe de vendas também é responsável por fornecer feedback e sugestões para o consultor sobre as funcionalidades e as melhorias do data warehouse, bem como por participar do treinamento e da adoção do data warehouse.

- **Gestor de planejamento de demanda:** é o funcionário da área de produção da Adventure Works, Luís Soares, que é responsável por planejar a produção dos produtos da empresa com base na demanda dos clientes. O gestor de planejamento de demanda é um potencial patrocinador de novos projetos de dados dentro da empresa, que gostaria de receber uma análise demonstrando o potencial de uma solução preditiva para melhorar o planejamento da produção. O gestor de planejamento de demanda é um usuário secundário do data warehouse de vendas, que deve utilizar os dados e as informações de vendas para alimentar a solução preditiva e comparar os resultados com o processo atual.

Etapas do Projeto

O projeto de data warehouse de vendas da Adventure Works foi dividido em cinco etapas principais:

- **Coleta de dados:** nesta etapa, realizou-se o levantamento de requisitos com as áreas de negócio e de TI, identificando as fontes, os tipos e os volumes de dados necessários para o projeto. Obteve-se o acesso aos dados por meio de arquivos em formato CSV. Utilizou-se a linguagem SQL para validar e explorar os dados, verificando a qualidade, a consistência e a relevância dos dados para o projeto.
- **Modelagem de dados:** nesta etapa, foi definido o modelo multidimensional de dados para o data warehouse de vendas. Identificou-se as medidas e as dimensões que compõem o data warehouse, bem como as relações entre elas, e também as granularidades, as hierarquias, as agregações e as métricas derivadas do data warehouse.
- **Construção de dados:** nesta etapa, foi implementado o ETL, utilizando a ferramenta dbt cloud, foi criado os scripts SQL para realizar as transformações necessárias nos dados, tais como limpeza, padronização, enriquecimento, validação, integração, etc. Também para realizar a carga dos dados no data warehouse, utilizando a ferramenta Google BigQuery. Foi utilizado o dbt cloud para executar, monitorar e testar o processo de ETL, bem como para documentar e versionar os scripts SQL. Antes de transformar os dados, foi necessário prepará-los. Isso envolveu a criação de tabelas de staging no DBT Cloud, que são versões brutas dos dados originais. Essas tabelas de staging servem como ponto de partida para as transformações subsequentes. Criação de tabelas intermediárias: Essas tabelas são usadas para realizar transformações complexas nos dados. Elas servem como um passo intermediário entre as tabelas de staging e as tabelas de fatos e dimensões. Já as tabelas de fatos contém as métricas principais do negócio,

enquanto as tabelas de dimensões contêm detalhes descritivos que permitem a análise dos dados de diferentes perspectivas. Teve também a configuração do arquivo `source.yml`, que define as fontes de dados para o DBT e como elas devem ser lidas. Ele foi configurado para apontar para as tabelas de staging corretas.

- **Implantação de dados:** nesta etapa, implantou-se o data warehouse de vendas no ambiente de produção, utilizando a ferramenta Google BigQuery. Foi utilizado o dbt cloud para implantar, monitorar e testar o data warehouse, bem como para documentar e versionar o código SQL.
- **Testes de Validação no DBT:** Os testes de validação são uma parte crucial do processo de transformação de dados no DBT. Eles garantem que os dados sejam precisos e confiáveis. Existem vários tipos de testes de validação que podem ser realizados, incluindo testes de unicidade, testes de não nulidade e testes de relacionamento. Os testes são definidos nos arquivos `.yml` e são executados usando o comando `dbt test`. Se um teste falhar, o DBT fornecerá informações detalhadas sobre o erro, permitindo que você identifique e corrija o problema. Os arquivos YAML (YAML Ain't Markup Language) são usados no DBT para definir modelos, testes e documentação. Eles são fáceis de ler e escrever, e permitem que você defina uma grande quantidade de configurações em um formato estruturado. No contexto do DBT, os arquivos YAML são usados para definir as fontes de dados (arquivo `source.yml`), os testes de validação e a documentação dos modelos. `DBT Test` e `DBT Run` são dois comandos fundamentais no DBT.
- **Análise de dados:** nesta etapa, realizou-se a análise dos dados do data warehouse de vendas, utilizando a ferramenta Power BI. Foi criado o relatório e dashboard para responder às perguntas de negócio levantadas pela diretoria da empresa, fornecendo insights e recomendações baseados em dados. A ferramenta Power BI foi utilizada para criar, visualizar e compartilhar os relatórios, dashboards e análises, bem como para documentar e versionar o código DAX.
- **Power BI e sua Interação com o Google BigQuery:** o Power BI é uma suíte de ferramentas de análise de negócios que permite visualizar seus dados e compartilhar insights em toda a organização. Ele se conecta a uma ampla variedade de fontes de dados, e é especialmente útil para criar relatórios interativos e painéis de controle. A interação do Power BI com o Google BigQuery é bastante direta. O Power BI tem um conector nativo para o BigQuery, o que significa que você pode importar dados diretamente do BigQuery para o Power BI. Uma vez que os dados são importados, você pode usar as ferramentas de visualização de dados do Power BI para criar relatórios e painéis.

Riscos/contingências

O projeto de data warehouse de vendas da Adventure Works está sujeito a alguns riscos que podem comprometer o seu sucesso. A seguir, são apresentados os principais riscos identificados, bem como as medidas de contingência planejadas para lidar com eles.

- **Resistência das áreas de negócio:** algumas áreas de negócio, como a área comercial, podem resistir à mudança e à adoção do data warehouse, por não verem o valor e o benefício do projeto, ou por preferirem investir em outras iniciativas. Esse risco pode afetar o engajamento, o feedback e a colaboração das áreas de negócio, bem como a utilização e a satisfação com o data warehouse. Para mitigar esse risco, deve-se envolver as áreas de negócio desde o início do projeto, comunicando os objetivos, os benefícios e os resultados esperados do data warehouse, bem como demonstrando casos de sucesso e melhores práticas de outras empresas que utilizam data warehouse. Também deve-se realizar treinamentos e workshops para capacitar e orientar as áreas de negócio na utilização do data warehouse, bem como coletar feedback e sugestões para melhorá-lo de acordo com as necessidades e as expectativas das áreas de negócio.
- **Limitação de recursos e de tempo:** o projeto pode enfrentar limitações de recursos humanos, financeiros e tecnológicos, que podem afetar o escopo, o cronograma e o orçamento do projeto. Além disso, o projeto pode enfrentar restrições de tempo, que podem afetar o prazo e a qualidade do projeto. Esse risco pode afetar o desempenho, a produtividade e a satisfação da equipe do projeto, bem como a entrega, a funcionalidade e a performance. Para mitigar esse risco, deve-se realizar um planejamento e uma priorização eficientes das atividades e das entregas do projeto, definindo o escopo, o cronograma e o orçamento do projeto, bem como os critérios de aceitação e de qualidade do projeto, também utilizar metodologias ágeis de gerenciamento de projetos.

Gerando Valor em Projetos de Infraestrutura de Dados

- **Melhorias na eficiência operacional:** o data warehouse de vendas pode ajudar a melhorar a eficiência operacional da empresa, reduzindo o tempo e o custo de acesso e de análise dos dados, automatizando e padronizando os processos de dados, eliminando redundâncias e inconsistências nos dados, e aumentando a confiabilidade e a segurança dos dados. Com isso, a empresa

pode economizar recursos, evitar erros e retrabalhos, e focar nas atividades de maior valor agregado.

- **Novas oportunidades de negócios:** o data warehouse de vendas pode ajudar a identificar novas oportunidades de negócios para a empresa, como novos produtos, novos mercados, novos canais, novos segmentos, novos parceiros, etc. Com isso, a empresa pode ampliar sua oferta, sua abrangência e sua competitividade no mercado de bicicletas, gerando mais receita e lucro.
- **Inovação e diferenciação:** o data warehouse de vendas pode ajudar a promover a inovação e a diferenciação da empresa, permitindo a criação de soluções avançadas de dados, como machine learning, inteligência artificial, big data, etc. Com isso, a empresa pode desenvolver produtos e serviços mais inteligentes, personalizados e adaptados às necessidades e às preferências dos clientes, criando uma vantagem competitiva e uma fidelização dos clientes.
- **Cultura data driven:** o data warehouse de vendas pode ajudar a criar e a fortalecer uma cultura data driven na empresa, incentivando o uso de dados para embasar as decisões, as ações e as estratégias da empresa, bem como para medir e melhorar os resultados e o desempenho da empresa. Com isso, a empresa pode se tornar mais ágil, mais orientada a resultados, mais alinhada aos objetivos e mais preparada para as mudanças e os desafios do mercado..