



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
DEPARTAMENTO DE AUTOMAÇÃO E SISTEMAS
CURSO DE GRADUAÇÃO EM ENGENHARIA DE CONTROLE E AUTOMAÇÃO

Leonardo Luiz Cubas

**Desenvolvimento de um sistema de business intelligence sob uma arquitetura
Data Science as a Service para uma empresa do ramo de manufatura de baterias
automotivas**

Florianópolis
2022

Leonardo Luiz Cubas

**Desenvolvimento de um sistema de business intelligence sob uma arquitetura
Data Science as a Service para uma empresa do ramo de manufatura de baterias
automotivas**

Relatório final da disciplina DAS5511 (Projeto de Fim de Curso) como Trabalho de Conclusão do Curso de Graduação em Engenharia de Controle e Automação da Universidade Federal de Santa Catarina em Florianópolis.

Orientador: Prof. Ricardo José Rabelo, Dr.

Supervisor: Igor Medeiros Benincá, Eng.

Florianópolis

2022

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Cubas, Leonardo Luiz

Desenvolvimento de um sistema de business intelligence sob uma arquitetura Data Science as a Service para uma empresa do ramo de manufatura de baterias automotivas / Leonardo Luiz Cubas ; orientador, Ricardo José Rabelo, coorientador, Igor Medeiros Benincá, 2022.

63 p.

Trabalho de Conclusão de Curso (graduação) - Universidade Federal de Santa Catarina, Centro Tecnológico, Graduação em Engenharia de Controle e Automação, Florianópolis, 2022.

Inclui referências.

1. Engenharia de Controle e Automação. 2. business intelligence. 3. data warehouse. 4. data science as a service. I. Rabelo, Ricardo José. II. Benincá, Igor Medeiros. III. Universidade Federal de Santa Catarina. Graduação em Engenharia de Controle e Automação. IV. Título.

Leonardo Luiz Cubas

Desenvolvimento de um sistema de business intelligence sob uma arquitetura Data Science as a Service para uma empresa do ramo de manufatura de baterias automotivas

Esta monografia foi julgada no contexto da disciplina DAS5511 (Projeto de Fim de Curso) e aprovada em sua forma final pelo Curso de Graduação em Engenharia de Controle e Automação

Florianópolis, 22 de julho de 2022.

Prof. Hector Bessa Silveira, Dr.
Coordenador do Curso

Banca Examinadora:

Prof. Ricardo José Rabelo, Dr.
Orientador
UFSC/CTC/DAS

Igor Medeiros Benincá, Eng.
Supervisor
Empresa Indicium Tech

Miguel Lino Ferreira Neto, Dr.
Avaliador
UFSC/CTC/DAS

Prof. Eduardo Camponogara, Dr.
Presidente da Banca
UFSC/CTC/DAS

Este trabalho é dedicado aos meus pais, às minhas
irmãs e ao meu supervisor na empresa.

RESUMO

As empresas de diferentes ramos têm cada vez mais optado por se tornar *data-driven*. Ou seja, elas buscam tomar decisões estratégicas baseadas em indicadores. Entretanto, para se atingir tal objetivo, a organização necessita, além de desenvolver uma maturidade analítica, implementar uma infraestrutura moderna e escalável de dados. Logo, a organização manufatora de baterias contratante do projeto descrito neste documento buscava ter acesso à painéis de *business intelligence* os quais contempassem dados comerciais da empresa, como volume vendido e faturamento de todas as unidades do país. Ao mesmo tempo buscava-se que a infraestrutura que suporta os relatórios fosse escalável e fizesse o uso de ferramentas adaptáveis ao crescente fluxo de informações. A solução proposta foi a modelagem e hospedagem de um *data warehouse* na nuvem, com ingestão de dados ocorrendo de forma diária, alimentando a ferramenta de *business intelligence*. Os dados dentro do repositório central foram formatados seguindo a metodologia *star schema*, com o relacionamento entre tabelas de fatos e dimensões. O fluxo de trabalho do projeto seguiu os conceitos da metodologia ágil *SCRUM*, com entregas recorrentes em *sprints* de duas semanas. Por fim, o resultado obtido foi uma arquitetura robusta de dados na nuvem, centralizada em um *data warehouse* formatado seguindo boas práticas de modelagem dimensional, que alimenta painéis de *business intelligence* desenvolvidos com o apoio de conceitos de *storytelling* e *UI (User Interface)*.

Palavras-chave: *Data-driven. Business intelligence. Data warehouse.*

ABSTRACT

Different field of work companies have decided to become data-driven. In other words, they seek to make strategic decisions based at performance indicators. However, the company needs to implement a scalable, modern data infrastructure to achieve this goal. Therefore, the battery manufacturing company approached at this document intended to develop business intelligence dashboards which showed comercial data from the company. At the same time, it wanted the data structure which supplies business intelligence reports to use scalable applications. The proposed solution was a cloud-hosted data warehouse implementation and modeling with daily data ingestion which supplies business intelligence tools. Data located at this central repository were formatted based on star-schema methodology. The project workflow was based on SCRUM agile methodology concepts. Finally, the result accomplished was a robust cloud-hosted data infrastructure centralized at a data warehouse, which was modeled based on dimensional modeling good practices. These data supply business intelligence reports developed based on storytelling and UI concepts.

Keywords: Data-driven. Business intelligence. Data warehouse.

LISTA DE FIGURAS

Figura 1 – Jornada <i>data driven</i> proposta pela Indicium Tech.	12
Figura 2 – Aba de modelos de dados no Power BI.	18
Figura 3 – Fluxo de boas práticas proposto pela <i>CRISP-DM</i>	20
Figura 4 – Exemplo de arquitetura de dados normalizada.	21
Figura 5 – Exemplo de metodologia <i>star-schema</i>	23
Figura 6 – Fluxo dos dados no dbt.	24
Figura 7 – Fluxo de atividades no <i>framework</i> SCRUM	25
Figura 8 – Vendas de baterias registradas no ERP geram dados no banco de dados operacional.	28
Figura 9 – Diagramas de implementação do <i>star-schema</i> dos dados do ERP AX.	34
Figura 10 – Fluxo dos dados no dbt.	35
Figura 11 – Fluxo de trabalho gitflow.	37
Figura 12 – Funcionamento dos diferentes ambientes do projeto e integração com as ramificações do gitflow.	38
Figura 13 – Fluxo dos dados no projeto.	40
Figura 14 – Organização da pasta <i>models</i> do dbt.	41
Figura 15 – Amostra de arquivos que compõem as pastas <i>staging</i> e <i>marts</i> no dbt.	42
Figura 16 – Etapas de transformação das tabelas do modelo dimensional no dbt.	43
Figura 17 – Amostra de diagrama gerado automaticamente na ferramenta dbt Docs.	45
Figura 18 – Esboço desenvolvido para proposta de narrativa de dados.	46
Figura 19 – Configuração das fontes de dados no Power BI.	47
Figura 20 – Parte 1 da aba geral do painel Mapeamento de Mercado.	49
Figura 21 – Parte 2 da aba geral do painel Mapeamento de Mercado.	49
Figura 22 – Parte 1 da aba de acompanhamento de meta de volume do painel Mapeamento de Mercado.	50
Figura 23 – Parte 2 da aba de acompanhamento de meta de volume do painel Mapeamento de Mercado.	50
Figura 24 – Parte 1 da aba de acompanhamento detalhado de volume do painel Mapeamento de Mercado.	51
Figura 25 – Parte 2 da aba de acompanhamento detalhado de volume do painel Mapeamento de Mercado.	52
Figura 26 – Parte 3 da aba de acompanhamento detalhado de volume do painel Mapeamento de Mercado.	53
Figura 27 – Seção de filtros de segmentação de dados.	54
Figura 28 – Conjunto de tecnologias englobadas em todas as frentes do projeto.	62

SUMÁRIO

1	INTRODUÇÃO	10
1.1	OBJETIVOS	10
1.2	CORRELAÇÃO COM O CURSO DE ENGENHARIA DE CONTROLE E AUTOMAÇÃO	11
1.3	ESTRUTURA DO DOCUMENTO	11
2	AS EMPRESAS	12
2.1	INDICIUM TECH: EMPRESA DE CONSULTORIA DE DADOS	12
2.2	A EMPRESA MANUFATORA DE BATERIAS	13
3	FUNDAMENTAÇÃO TEÓRICA	14
3.1	INDICADORES DE MERCADO NA TOMADA DE DECISÕES	14
3.2	<i>BUSINESS INTELLIGENCE</i>	16
3.3	<i>POWER BI</i>	17
3.4	<i>CROSS-INDUSTRY PROCESS FOR DATA-MINING (CRISP-DM)</i>	18
3.5	TRANSFORMAÇÃO DE DADOS	19
3.5.1	Modelagem dimensional e metodologia <i>star schema</i>	20
3.6	O <i>DBT</i> : FERRAMENTA DE TRANSFORMAÇÃO DE DADOS	22
3.7	SCRUM: METODOLOGIA ÁGIL PARA GERENCIAMENTO DE PROJETOS	24
4	ESPECIFICAÇÃO DO PROJETO	27
4.1	ENTENDIMENTO DE NEGÓCIO	27
4.1.1	Fluxo antigo de dados do setor comercial	27
4.1.2	Escopo do projeto	27
4.1.3	Usuários de negócio	28
4.1.4	Requisitos do projeto	28
4.1.4.1	Requisitos funcionais	29
4.1.4.2	Requisitos não funcionais	29
4.1.5	Critérios de sucesso	30
4.2	ENTENDIMENTO DOS DADOS	31
4.2.1	Material disponibilizado	31
4.2.2	Mapeamento do painel de mapeamento de mercado	31
4.2.3	Indicadores mapeados	32
4.2.4	Criação de documentação e modelos de dados	32
4.3	FLUXO DE MODELAGEM DE DADOS	33
4.3.1	Área <i>staging</i>	34
4.3.1.1	Renomeação de colunas	34
4.3.1.2	Mudança de tipos de dados	35
4.3.1.3	Tratamento de dados	35

4.3.1.4	Seleção de colunas	35
4.3.2	Área <i>marts</i>	36
4.3.3	Versionamento de código	36
4.3.4	Ambientes de desenvolvimento, validação e produção	37
4.3.4.1	Ambiente de desenvolvimento	38
4.3.4.2	Ambiente de validação	39
4.3.4.3	Ambiente de produção	39
5	IMPLEMENTAÇÃO DO PROJETO	40
5.1	TRANSFORMAÇÃO DOS DADOS	40
5.1.1	Organização do repositório do dbt	41
5.2	CRIAÇÃO DE TESTES DE DADOS E DOCUMENTAÇÃO	42
5.2.1	Testes de dados	42
5.2.2	Documentação da transformação de dados	44
5.2.2.1	dbt Docs	44
5.3	DESENVOLVIMENTO DO PAINEL DE <i>BUSINESS INTELLIGENCE</i>	44
5.3.1	Criação de esboço do painel	45
5.3.2	Configuração do Power BI	46
5.3.2.1	Modelo de dados	46
5.3.2.2	Conexão de dados	47
5.3.3	Desenvolvimento da interface de navegação	48
5.3.4	Implementação dos visuais	48
5.3.4.1	Aba geral	48
5.3.4.2	Aba de acompanhamento da meta de volume	49
5.3.4.3	Aba de acompanhamento detalhado de volume	51
5.3.4.4	Seção de filtros	51
6	RESULTADOS FINAIS	55
6.1	ATENDIMENTO AOS REQUISITOS E CRITÉRIOS DE SUCESSO	55
6.1.1	Requisitos funcionais	55
6.1.2	Requisitos não funcionais	55
6.1.3	Critérios de sucesso	56
6.2	VALIDAÇÃO DOS DADOS	56
7	CONCLUSÃO	58
7.1	PROPOSTA DE TRABALHOS FUTUROS	58
	REFERÊNCIAS	59
	ANEXO A – TECNOLOGIAS UTILIZADAS	62

1 INTRODUÇÃO

Com o crescente barateamento dos custos de armazenamento de dados e computação na nuvem, muitas empresas têm optado por se tornar *data-driven*, ou seja, tomar decisões estratégicas baseadas em dados. Entretanto, é comum que as mesmas não possuam uma infraestrutura de dados adaptada para fornecer indicadores estratégicos, além de carecer de colaboradores com conhecimento técnico para implementação da mesma. Nesse contexto, tem-se recorrido a empresas terceirizadas visando a implementação de tecnologias que proporcionam a análise de dados da organização. Tal processo é conhecido como *Data Science as a Service* (DSaaS).

No projeto descrito neste documento, uma empresa do ramo de manufatura de baterias busca justamente implementar uma infraestrutura de dados escalável que suporte a criação de relatórios de *business intelligence* com o intuito de possibilitar tomadas de decisão baseadas em dados.

Para o desenvolvimento do projeto, fez-se o uso de algumas metodologias. No contexto da gestão e planejamento do desenvolvimento, adotou-se a metodologia ágil SCRUM, na qual divide-se o tempo de projeto em ciclos chamados de *sprints*, com reuniões diárias de alinhamento. Com o intuito de organizar o fluxo de desenvolvimento do projeto de dados em etapas, utilizou-se a metodologia CRISP-DM (*cross-industry process for data-mining*). Por fim, foram utilizados conceitos de modelagem dimensional com o objetivo de estruturar os dados em formatos adaptados para abastecimento de painéis de *business intelligence*.

O papel do autor no desenvolvimento do projeto foi baseado no mapeamento das fontes de dados da empresa manufadora de baterias, transformação dos dados seguindo conceitos e boas práticas de modelagem dimensional e construção dos painéis de *business intelligence* performáticos.

1.1 OBJETIVOS

Busca-se, na implementação do projeto descrito nesta monografia, implementar uma infraestrutura de dados escalável e otimizada para o abastecimento de uma ferramenta de *business intelligence*. A seguir, são descritos objetivos específicos do projeto:

1. Entrega de um *data warehouse* hospedado em ferramenta de computação na nuvem, o qual é modelado de forma a entregar tabelas tratadas e otimizadas para abastecimento de ferramentas de *business intelligence*;
2. Desenvolvimento de relatório de *business intelligence* que apresenta indicadores estratégicos da área comercial da empresa contratante;

3. Elaboração de documentação do projeto cuja leitura permita realização de eventuais manutenções e replicação do projeto para outras áreas da empresa;
4. Disponibilização de dados confiáveis aos usuários de negócio.

1.2 CORRELAÇÃO COM O CURSO DE ENGENHARIA DE CONTROLE E AUTOMAÇÃO

O presente projeto se correlaciona com o curso de Engenharia de Controle e Automação ao aplicar conceitos vistos nas disciplinas de metodologia para desenvolvimento de processos, avaliação de desempenho de sistemas de automação discretas, integração de sistemas corporativos, *data warehouse* e sistemas inteligentes.

O raciocínio crítico e analítico desenvolvido pelo autor com o curso de tais disciplinas também foi fundamental em diversas etapas da implementação do projeto.

1.3 ESTRUTURA DO DOCUMENTO

Esta monografia está dividida em 7 capítulos principais. Nos capítulos 1 e 2 descreve-se a introdução ao problema, os objetivos do projeto, a correlação do mesmo com o curso de Engenharia de Controle e Automação e o contexto básico da empresa contratada e contratante.

Em seguida, no capítulo 3, descreve-se os conceitos, metodologias e ferramentas que basearam a implementação da solução do projeto. Já nos capítulos 4, 5 e 6 são abordadas as especificações técnicas do projeto, o processo de implementação dos conceitos previamente apresentados e os resultados finais.

Por fim, o capítulo 7 engloba as conclusões do autor e sugestões de possíveis trabalhos futuros.

2 AS EMPRESAS

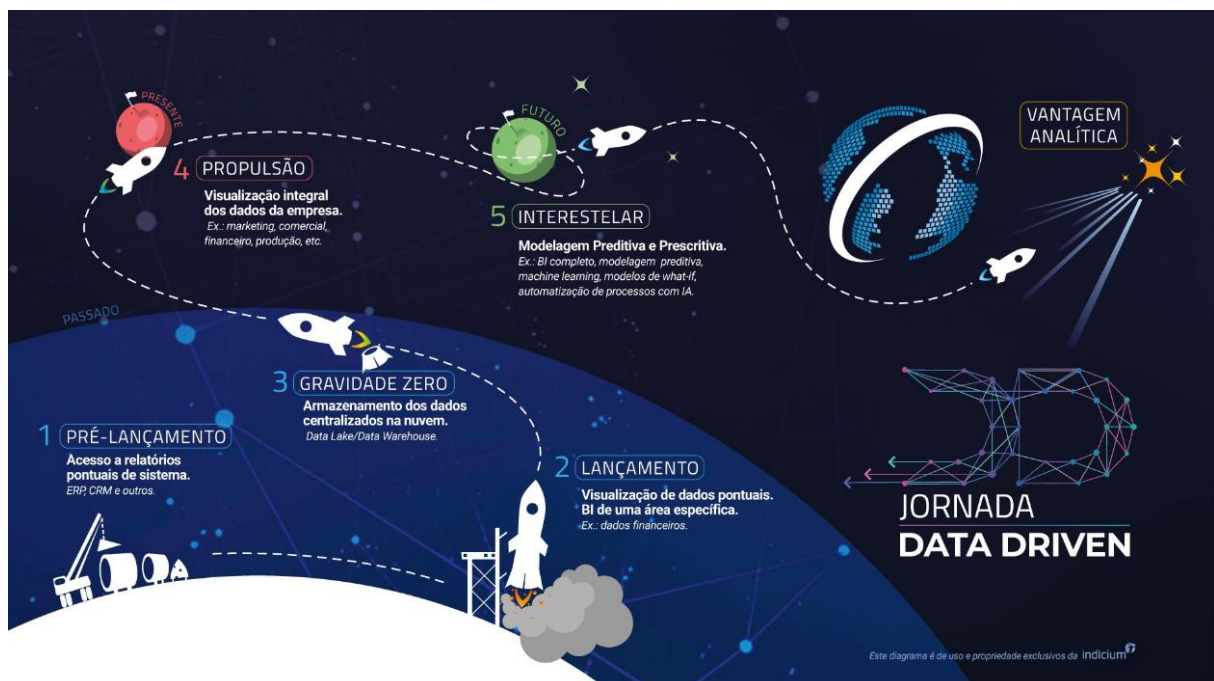
2.1 INDICIUM TECH: EMPRESA DE CONSULTORIA DE DADOS

Sediada atualmente no norte da Ilha de Santa Catarina, em Florianópolis, a Indicium Tech conta com cerca de 100 colaboradores. O trabalho pode ser feito tanto de forma *home office* como presencial.

A Indicium Tech é uma empresa que presta consultoria de dados a outras grandes companhias que desejam tomar decisões baseadas em análises de dados. Logo, não vende produtos de *software* prontos, mas desenvolve ferramentas sob demanda para cada cliente no formato *Data Science as a Service*. A companhia está envolvida em projetos de empresas dos setores de indústria, logística, energia, *marketing*, dentre outros.

Uma das estratégias desenvolvidas pela Indicium Tech para atrair novos clientes e organizar o processo de implementação de inteligência de dados é a jornada *data-driven*. Como pode-se observar na figura 1, tal processo define em qual nível de amadurecimento de dados a empresa contratante dos seus serviços está, contribuindo para a definição do escopo do projeto.

Figura 1 – Jornada *data driven* proposta pela Indicium Tech.



Fonte: (BLASI, 2021).

2.2 A EMPRESA MANUFATORA DE BATERIAS

Sediada no Estado de Pernambuco, a empresa manufatora de baterias elétricas solicitante do projeto em questão possui mais de 80 distribuidores no Brasil, Argentina, Uruguai e Paraguai. Ela tem atuação em diversos mercados, como automotivo, náutico, energia alternativa, dentre outros.

Desde a década passada, a empresa tem investido na implementação de uma infraestrutura de dados robusta com o objetivo de possibilitar a criação de indicadores que suportem tomadas de decisões estratégicas baseada em dados. Entretanto, a mesma ainda carece de um repositório central definido como fonte única de verdade para a organização.

Optou-se pela ocultação do nome da empresa neste documento com o intuito de preservar a inteligência de negócio e informações referentes à arquitetura de dados da mesma.

3 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo são descritos os conceitos necessários para o pleno entendimento do contexto em que este projeto está envolvido, assim como metodologias e ferramentas que auxiliaram o autor no desenvolvimento do mesmo.

3.1 INDICADORES DE MERCADO NA TOMADA DE DECISÕES

As empresas dos mais diversos ramos têm optado por tomar decisões estratégicas baseadas em dados. Entretanto, há uma grande variedade de informações que podem ser relevantes para a escolha de um planejamento estratégico que resulte em benefícios para a organização. Assim, a compreensão e criação dos indicadores de performance (do inglês *Key Performance Indicators*, ou *KPIs*) é uma ótima estratégia de resolução.

Os *KPIs* são métricas monitoradas com o objetivo de determinar a saúde, eficiência e eficácia de uma organização. Eles são medidas quantitativas que auxiliam os usuários envolvidos a entender se a companhia está cumprindo seus objetivos, e possuem definição única e clara dentro de uma organização. Os indicadores de performance são um dos pilares das tomadas de decisões estratégicas, as quais têm como objetivo alinhar a empresa com seu ambiente em constante mudança (DOLENCE; NORRIS, 1994).

Indicadores de performance podem ser classificados em onze grupos de acordo com a sua aplicação e seus benefícios em uma empresa (BHATTI; AWAN; RAZAQ, 2014):

1. **Qualidade:** garantir a qualidade do produto final é primordial para empresas que querem se destacar no cenário nacional e internacional nos dias atuais. Indicadores podem auxiliar neste sentido ao informar o usuário dados referentes à confiabilidade de máquinas industriais, quantidade de produtos defeituosos e durabilidade dos mesmos;
2. **Flexibilidade:** a flexibilidade de uma empresa determina o quão maleável a mesma é com o objetivo de atender diferentes demandas comerciais, garantindo uma expansão na gama de clientes. Tempos médios de troca de maquinários para produzir um item diferente, ou mesmo a quantidade de produtos distintos que a empresa manufatura são *KPIs* encontrados no mercado;
3. **Tempo:** indicadores relacionados ao tempo são usados em larga escala em organizações de diversos setores. Mapear o tempo gasto em determinados processos possibilita que a empresa tome ações com o objetivo de tornar o setor produtivo mais eficiente, por exemplo. Taxa de produção, tempo de espera e frequência

de venda de produtos são alguns dos exemplos de indicadores baseados na dimensão tempo;

4. Segurança: o mapeamento de indicadores relacionados à segurança de funcionários e clientes de uma empresa tem se tornado comum nas organizações, principalmente no setor industrial. Taxas de acidentes e frequência de treinamentos de segurança são alguns dos indicadores utilizados por tomadores de decisão para acompanhar a situação da organização neste viés;
5. Performance financeira: o cálculo de indicadores de performance financeira é crucial para a avaliação do desempenho de uma empresa. Com tais métricas é possível acompanhar o impacto de campanhas de *marketing*, destinar orçamento a setores mais atrasados tecnologicamente e verificar o crescimento econômico da empresa. Alguns dos *KPIs* mais utilizados para medir performance financeira são: volume vendido, fluxo de caixa, faturamento líquido e crescimento nas vendas;
6. Custo: o mapeamento de custos se mostra importante ao informar os gerentes se os gastos com mão de obra, serviços e matéria prima estão dentro da margem planejada. Custo médio de fabricação de um produto, gastos com funcionários e serviços e taxas de produtividades são *KPIs* comumente encontrados em relatórios de tomadores de decisões;
7. Satisfação de funcionários: funcionários satisfeitos com seus cargos tendem a ser mais performáticos no ambiente de trabalho. A satisfação pode ser avaliada em cinco dimensões: remuneração, qualidade da rotina de trabalho, promoção de cargo, supervisão e trabalho em equipe (OMAH; OBIEKWE, 2019). Indicadores empregados com frequência nas empresas são: quantidade de funcionários satisfeitos em pesquisas internas e rotatividade de pessoal;
8. Aprendizado e crescimento: o constante desenvolvimento de seus funcionários é um diferencial para empresas que buscam expansão de mercado. É inviável uma empresa produzir resultados efetivos e distintivos sem o trabalho extraordinário de seus funcionários (FAROOQ; KHAN, 2011). Quantidade de promoções internas, horas gastas em treinamento e investimento financeiro em programas de capacitação são alguns dos indicadores mais utilizados nesta dimensão;
9. Performance social: empresas têm um grande impacto na sociedade no geral. Indicadores como porcentagem da população local empregada, quantidade de campanhas sociais promovidas e desperdício produzido são amplamente encontrados nas organizações;

10. Satisfação de clientes: a satisfação de clientes tem impacto direto na performance financeira de uma empresa, pois consumidores satisfeitos tendem a repetir transações comerciais com a mesma, além de promover a imagem da organização através de *buzz marketing* (conhecido popularmente como *marketing boca a boca*). Quantidade de reclamações, frequência de pedidos por cliente e contagem de novos clientes são alguns dos *KPIs* utilizados em relatórios pelo mundo;
11. Confiança em entregas: é comum organizações pelo mundo não terem acesso a indicadores que informem se seus respectivos produtos finais satisfazem o nível de exigência adotado pelos gerentes. Tempo médio de atraso, porcentagem de produtos vendidos devolvidos em decorrência de defeitos técnicos são indicadores visados neste sentido.

3.2 BUSINESS INTELLIGENCE

Business intelligence é um termo amplo designado para um grupo de conceitos e tecnologias implementadas em conjunto com o objetivo de auxiliar executivos, gerentes e analistas a tomar melhores e mais rápidas decisões (CHAUDHURI; DAYAL; NARASAYYA, 2011).

O conceito de *BI* engloba não somente os *softwares* onde os dados são visualizados, mas também o conjunto de boas práticas de formatação dos mesmos e toda a infraestrutura que suporta o constante abastecimento de informações (GARTNER.COM, 2022).

Com o objetivo de ter seus dados visualizados de forma integrada e confiável em *softwares* de *business intelligence*, as empresas devem antes investir na implementação de uma arquitetura de dados escalável e que permita o cruzamento de dados de diferentes fontes. Tal arquitetura está dividida nas seguintes etapas:

1. Bancos de dados transacionais: grande parte dos *softwares* utilizados em organizações para gerir seus processos funcionam com um processamento de dados em tempo real (do inglês *Online Transaction Processing*, ou *OLTP*). Dados de entradas e saídas de pedidos de compras e transações bancárias são exemplos de informações armazenadas em sistemas *OLTP*. Os dados produzidos por estes sistemas transacionais são posteriormente tratados para visualização em um formato de suporte à tomada de decisões estratégicas;
2. Ingestão de dados: o processo de extração e carga dos dados de sistemas *OLTP* para um sistema de armazenamento central pode acontecer de duas formas de acordo com o sequenciamento das operações de ingestão e tratamento dos dados:

- a) *ETL (Extract, Transform, Load)*: no processo de *ETL* (em português Extração, Transformação, Carga) os dados são extraídos dos bancos de dados operacionais e, em sequência, são transformados para um formato mais propício para aplicação em ferramentas analíticas. Por fim, os dados são carregados em um armazenamento de dados central (o *data warehouse*);
 - b) *ELT (Extract, Load, Transform)*: já o processo de *ELT* (em português Extração, Carga, Transformação) é uma adaptação do *ETL*, no qual os dados são carregados no *data warehouse* sem transformações prévias, antes de serem formatados.
3. Armazenamento central: é desejável nas empresas ter um repositório central como fonte única de verdade. Segundo Ralph Kimball em seu livro *The Data Warehouse Toolkit* (KIMBALL; ROSS, 2013) a implementação de um *data warehouse* objetiva tornar a informação facilmente acessível a todos usuários de negócio, construir uma fonte confiável para tomadas de decisões e ter um local com informações consistentes e facilmente compreendidas, dentre outras premissas;
 4. Visualização: a leitura de relatórios que apresentem os principais indicadores da empresa é, geralmente, o objetivo final em um projeto de *business intelligence*. Portanto, os *softwares* de BI disponibilizam um ferramental variado para a criação de gráficos e tabelas que apresentem os KPIs em formatos desejados pelo usuário final. Filtros podem ser adicionados nos relatórios para que o tomador de decisão interaja com os dados apresentados.

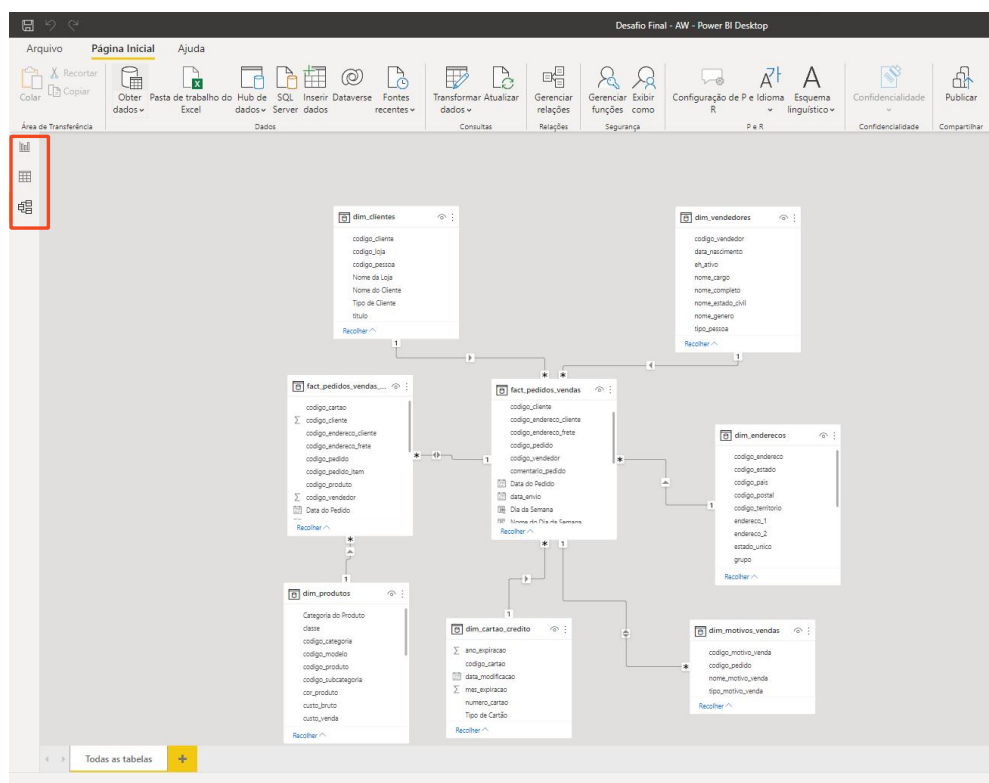
3.3 POWER BI

O *Power BI* é uma ferramenta desenvolvida pela *Microsoft* que engloba uma coleção de tecnologias, aplicativos e conectores com o objetivo de possibilitar que seu usuário construa relatórios integrados para tomadas de decisões. O *software* possui versões para *desktop*, navegador e dispositivos móveis (MICROSOFT.COM, 2022).

A versão utilizada para desenvolvimento dos painéis neste projeto foi a *desktop*, a qual é segmentada em três seções com funcionalidades específicas:

1. Relatório: setor do aplicativo onde os painéis são desenvolvidos;
2. Dados: local onde dados de tabelas podem ser consultados, além de possibilitar a criação de novas tabelas, colunas e métricas;
3. Modelo: seção na qual é criado o relacionamento entre as tabelas, o qual afetará diretamente o comportamento dos gráficos, filtros e tabelas do relatório. A seção pode ser visualizada na figura 2.

Figura 2 – Aba de modelos de dados no Power BI.



Fonte: autoria própria

No projeto descrito neste documento, todas as seções do *Power BI* foram usufruídas para a construção dos painéis de indicadores. Enquanto a seção Relatório foi a mais utilizada por abrigar as ferramentas de criação de visuais de indicadores, a seção Modelo foi constantemente atualizado com o objetivo de fornecer um cruzamento perfeito entre os dados das diferentes tabelas.

3.4 CROSS-INDUSTRY PROCESS FOR DATA-MINING (CRISP-DM)

A *Cross-industry process for data-mining* (CRISP-DM) é uma metodologia a qual apresenta um fluxo de desenvolvimento de projetos na área da ciência de dados e *analytics*. Ela está dividida em seis etapas (SCHRÖER; KRUSE; GÓMEZ, 2021):

1. Entendimento do negócio: escolha dos objetivos da empresa com o desenvolvimento do projeto, determinação de quais recursos tecnológicos a mesma já

possui e quais devem ser implementados e criação de um planejamento detalhado;

2. Entendimento dos dados: aquisição de dados para exploração dos mesmos, entendimento dos seus significados e como se relacionam entre si, teste de qualidade para verificar eventuais anomalias e criação de uma documentação explanando os pontos previamente citados;
3. Transformação dos dados: seleção das informações que serão utilizadas no modelo, tratamento de dados anômalos, formatação de acordo com o objetivo de negócio e integração de informações de fontes diversas;
4. Modelagem: seleção da técnica de treinamento do modelo, implementação do caso de teste e do modelo final;
5. Avaliação: revisão dos resultados, comparação dos mesmos com os objetivos propostos na primeira etapa e decisões de ações baseadas no que se foi atingido;
6. Implantação: planejamento do *deployment* do projeto, efetiva implantação, monitoramento e manutenção da solução ou relatório gerado.

A metodologia CRISP-DM norteia o planejamento de projetos nas mais diversas áreas de dados e deve ser adaptada para cada caso. No projeto descrito nesta monografia a fase de modelagem foi adaptada para o desenvolvimento de relatórios de *business intelligence*.

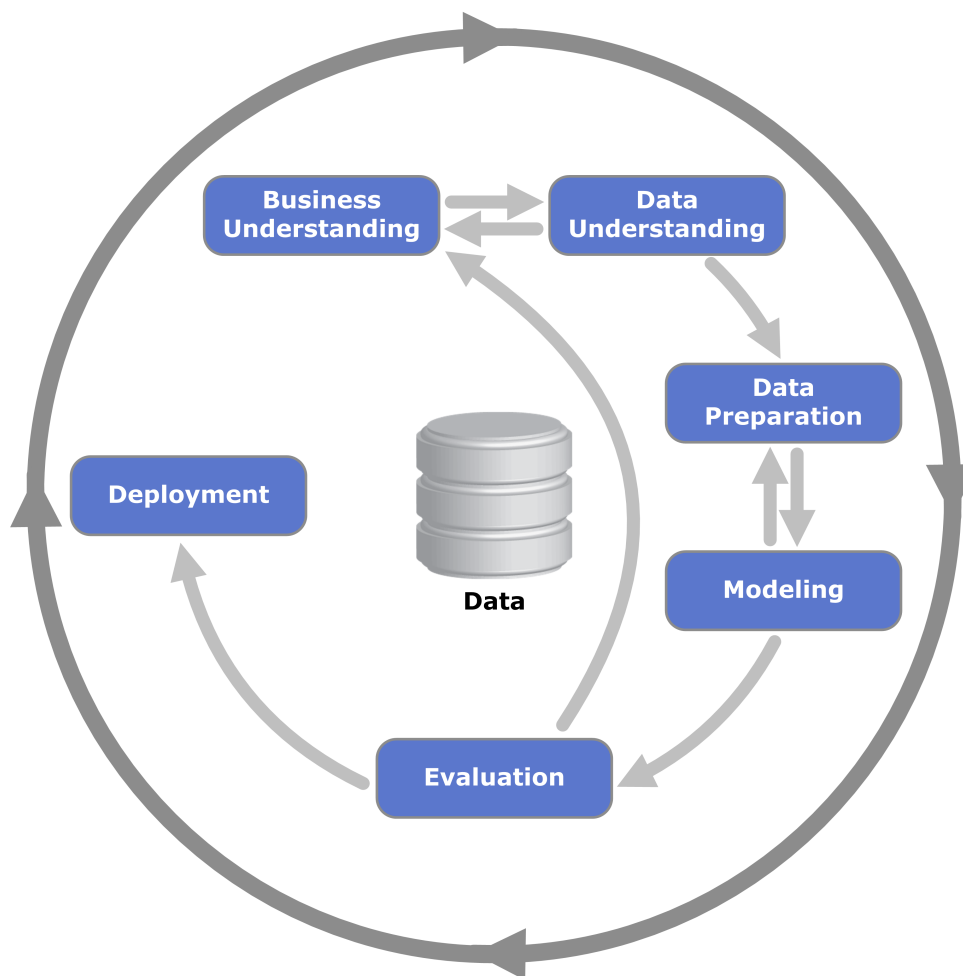
3.5 TRANSFORMAÇÃO DE DADOS

Devido ao crescente barateamento dos custos de geração e armazenamento de dados com o avanço de tecnologias de ETL e computação em nuvem, as organizações têm optado por incluir em seus planejamentos estratégicos maiores investimentos na geração de dados e criação de informações. Todavia, esses dados são comumente armazenados em formatos transacionais a partir de sistemas de processamento OLTP, o qual registra eventos do processo da empresa.

Além disso, grande parte dos dados gerados não têm utilidade analítica para os tomadores de decisão da empresa. Ou seja, eles não agregam valor à tomada de decisão estratégica da organização.

Por fim, parte dos dados costuma apresentar valores anômalos que podem diminuir a confiabilidade dos indicadores estratégicos.

A transformação dos dados visa justamente entregar informações confiáveis que auxiliarão na tomada de decisões da empresa, além de, frequentemente, possibilitar o cruzamento de dados de diferentes fontes, algo que é imprescindível para a implementação de um *data warehouse*. (RAHM; DO, 2000)

Figura 3 – Fluxo de boas práticas proposto pela *CRISP-DM*.

Fonte: (JENSEN, 2012)

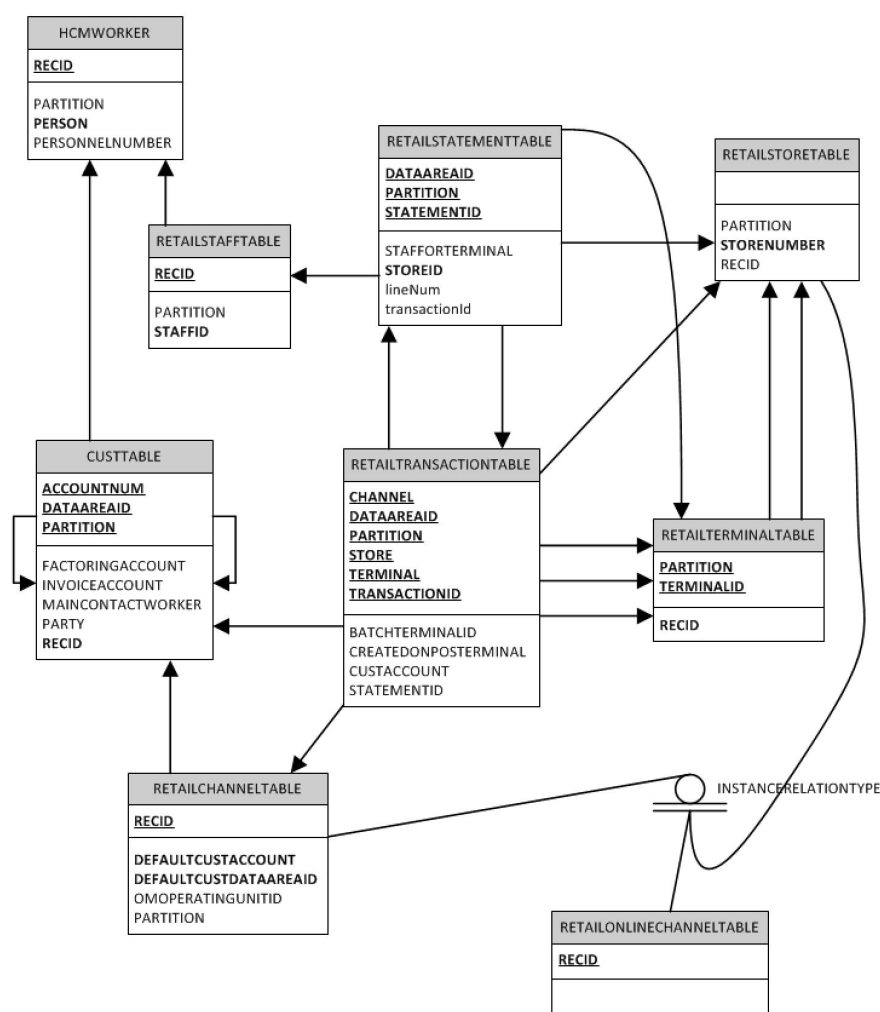
3.5.1 Modelagem dimensional e metodologia *star schema*

Modelagem dimensional é uma técnica de formatação de dados voltada para visualizações analíticas dos mesmos. Os principais objetivos são disponibilizar informações entendíveis para o usuário de negócio e aumentar a performance de consultas. A simplicidade na formatação dos dados auxilia tomadores de decisão, os quais com frequência carecem de conhecimentos mais técnicos de bancos de dados, a visualizar os processos da empresa em um modelo de informações enxuto, no qual suas respectivas tabelas e colunas são renomeadas a partir de uma linguagem humanizada (KIMBALL; ROSS, 2013).

A modelagem dimensional tem um papel ímpar na adaptação de modelos de dados normalizados (conhecidos como modelos 3NF, ou *third normal form*) provenientes de sistemas de dados OLTP. Tais estruturas são utilizadas em arquiteturas operacionais, pois cada transação da empresa impacta apenas uma tabela. Entretanto, tal formatação gera inúmeros relacionamentos entre tabelas, tornando o modelo complexo e mal otimizado para inserção em ferramentas de *business intelligence*. Os

relacionamentos das tabelas formatadas sob uma arquitetura 3NF são visualizados, com frequência, em diagramas ER (*entity-relationship diagrams*), conforme visto na figura 4.

Figura 4 – Exemplo de arquitetura de dados normalizada.



Fonte: alexdmeyer.com (MEYER, 2003)

Os relacionamentos entre tabelas de fatos e dimensões, as quais estão descritas nos tópicos a seguir, são definidos pela modelagem dimensional.

1. Tabelas de fatos: armazenam métricas relacionadas aos eventos dos processos

da empresa. Todas as linhas de uma tabela de fatos devem estar na mesma granularidade, ou seja, em uma mesma ordem de grandeza. Por exemplo, em uma tabela de vendas cada linha deve ser atribuída à um evento de venda apenas. Tal fator garante que métricas não contabilizem valores duplicados. Elas costumam possuir uma grande quantidade de linhas e ocupar cerca de 90% do espaço de armazenamento total dos modelos. Tabelas de fatos possuem, com frequência, duas ou mais chaves estrangeiras, as quais se relacionam com as chaves primárias das tabelas de dimensões;

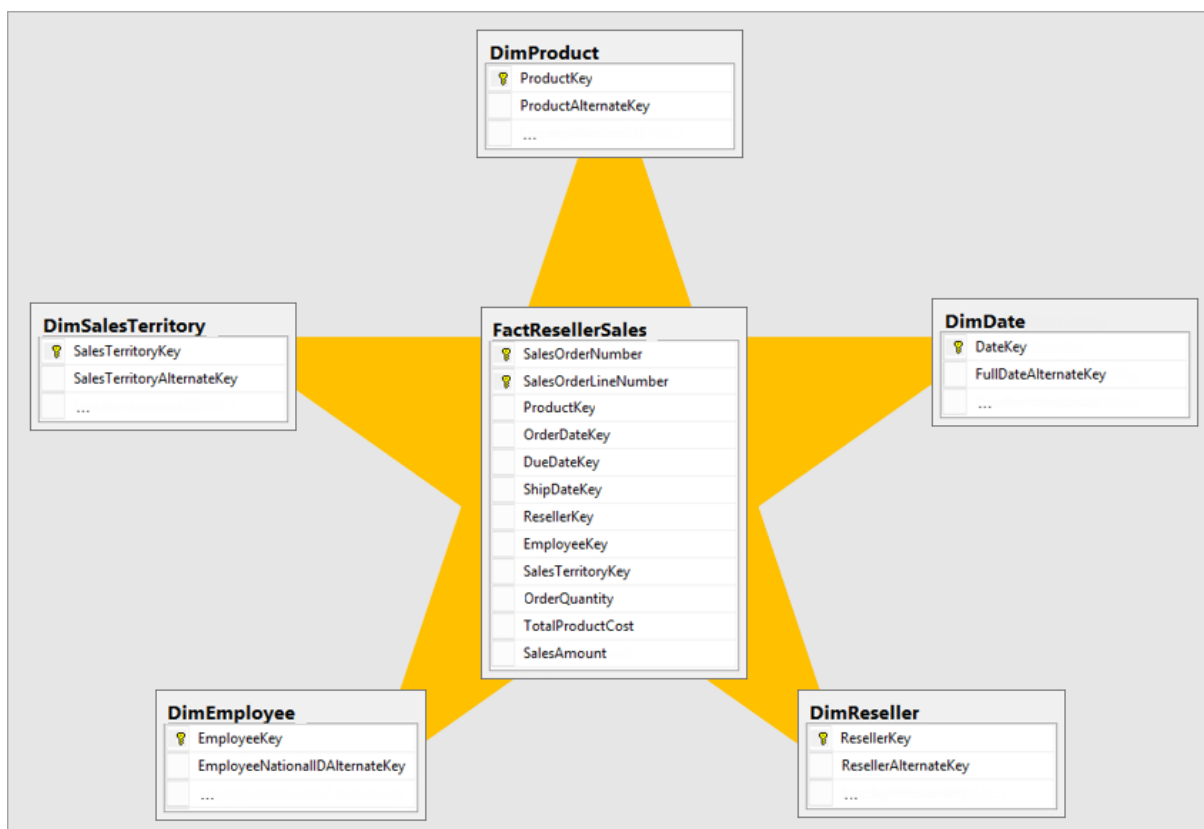
2. Tabelas de dimensões: armazenam dados descritivos os quais servem de contexto às tabelas de fatos, ou seja, descrevem o “o que, quem, onde, quando, como e por quê” associados aos eventos (KIMBALL; ROSS, 2013). As dimensões costumam atuar como restrições às visualizações e consultas de indicadores, também sendo parâmetro para agrupamentos de dados. Clientes, produtos, datas, vendedores, dentre outros, costumam nortear as tabelas de dimensões, as quais são compostas por atributos dessas entidades. Por exemplo, uma tabela de dimensões de clientes de uma empresa agrupa todas as informações dos mesmos e suas respectivas colunas são utilizadas para contextualizar as métricas das tabelas de fatos nos painéis de *business intelligence*.

O relacionamento entre tabelas de fatos e dimensões é descrito através da metodologia *star-schema* (esquema em estrela em português), na qual é referenciada uma tabela de fatos apenas e suas respectivas tabelas de dimensões interligadas através de chaves primárias e estrangeiras, como pode ser observado na figura 5. Um único projeto de implementação de painéis de *business intelligence* pode possuir diversos esquemas em estrela. Entretanto, deve-se evitar o relacionamento entre tabelas de fatos e dimensões entre si, já que tal interligação aumenta a complexidade do modelo de dados e descaracteriza um *star-schema*. Modelos que possuem relacionamentos entre tabelas de dimensões são conhecidos como *snowflake*.

Por fim, pode-se concluir que a adaptação de modelos normalizados para modelos dimensionais é imprescindível em um projeto de *business intelligence*. O relacionamento entre as tabelas de fatos e dimensões seguindo a metodologia *star-schema* possibilita que o usuário de negócio visualize no modelo de dados o processo da empresa de forma mais intuitiva.

3.6 O DBT: FERRAMENTA DE TRANSFORMAÇÃO DE DADOS

O dbt é uma ferramenta de transformação de dados a qual permite que seus usuários formatem as informações em um *data warehouse* através de *scripts* na linguagem SQL (linguagem padrão de consulta de bancos de dados). A ferramenta se destaca, também, por possibilitar a implementação de boas práticas de versionamento

Figura 5 – Exemplo de metodologia *star-schema*.

Fonte: docs.microsoft.com (MICROSOFT, 2022)

de código, desenvolvimento de documentação e criação de testes de robustez de dados.

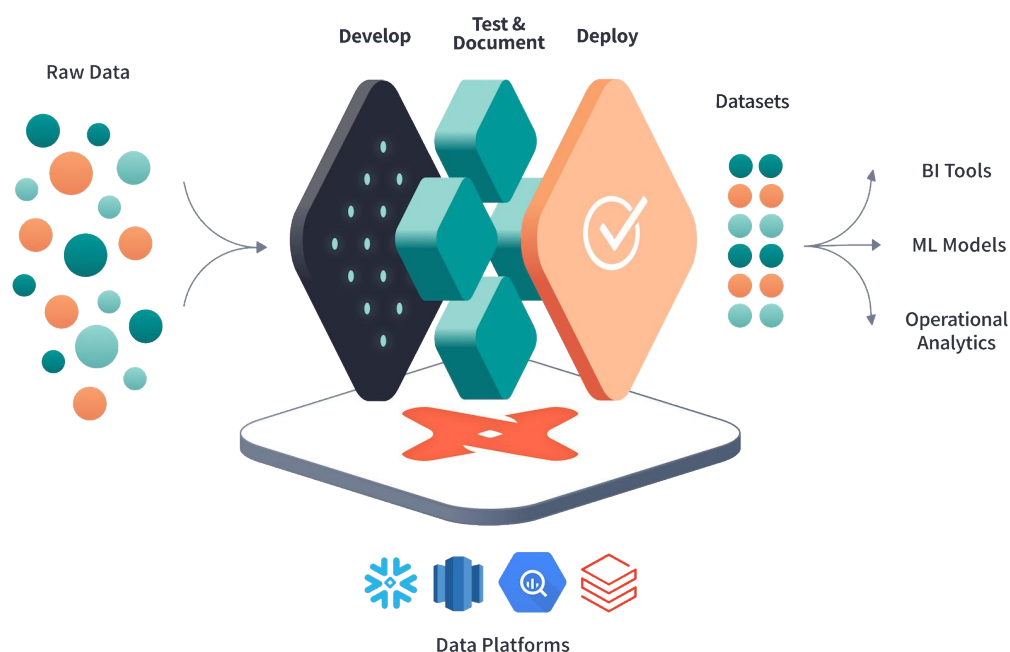
A aplicação possui uma versão disponível na nuvem, o dbt Cloud, e outra *open-source* em linhas de comando, o dbt Core. Enquanto a primeira possui planos grátis e pagos de acordo com a disponibilidade de funcionalidades, a segunda é totalmente gratuita.

O dbt possibilita que transformações de dados sejam realizadas com conhecimento básico da linguagem padrão de bancos de dados (SQL), logo não dependendo de engenheiros de dados com conhecimentos mais técnicos em determinadas linguagens de programação. Os próprios analistas de dados conseguem, com a ferramenta, tratar dados para consumo em relatórios de *business intelligence*. A modularidade do repositório aliada ao intuitivo versionamento de código auxilia no desenvolvimento de projetos colaborativos.

Por fim, a aplicação possibilita que o usuário realize a implementação das transformações de dados, testes de dados e documentação em um ambiente intermediário de teste. Tal fator é preponderante para que mudanças nos modelos apenas sejam notadas pelos usuários finais dos dados após o *deployment* das mesmas. Há a possibi-

lidade de criação de diversos outros ambientes (nomeados como perfis na ferramenta), os quais geralmente configuram um local de produção, o qual fornece dados já validados e testados para o usuário final. O fluxo dos dados dentro do dbt pode ser visualizado na figura 6.

Figura 6 – Fluxo dos dados no dbt.



Fonte: getdbt.com (LABS, 2022)

No projeto retratado nesta monografia a ferramenta dbt teve papel fundamental ao auxiliar o presente autor na estruturação dos dados do setor comercial da empresa manufatora de baterias seguindo conceitos de modelagem dimensional e da metodologia *star-schema*.

3.7 SCRUM: METODOLOGIA ÁGIL PARA GERENCIAMENTO DE PROJETOS

Na área da tecnologia da informação existem diversas metodologias de desenvolvimento de sistemas (*systems development methodologies*, ou SDM) que visam aumentar a produtividade das equipes através de processos e fluxos de trabalho. Elas podem ser divididas em tradicionais e ágeis (CHAN; THONG, 2009).

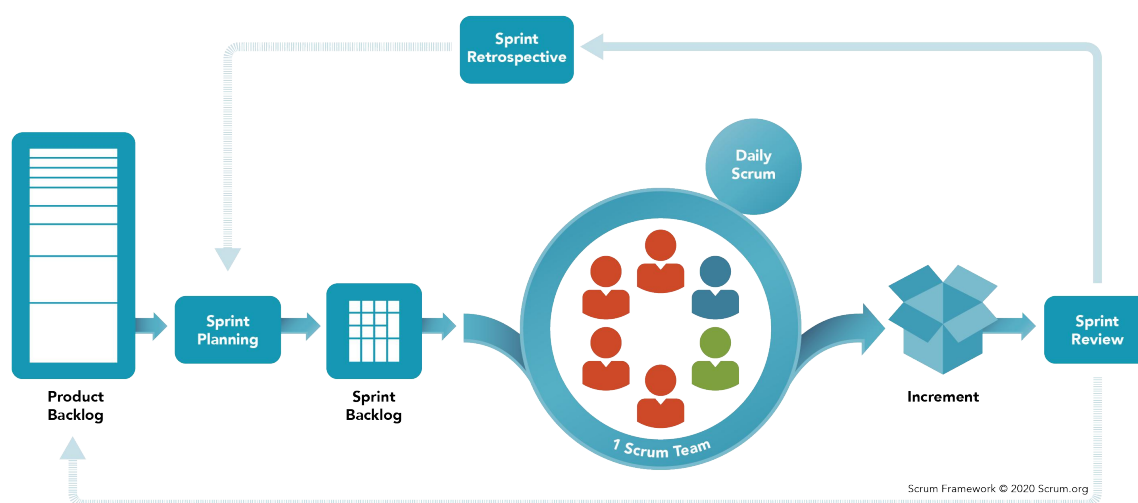
As metodologias de desenvolvimento de sistemas tradicionais se baseiam em um fluxo linear de fases preestabelecidas de trabalho. Esta abordagem é centrada no processo, e tem como premissa que eventuais fontes de variação do planejamento são identificáveis e podem ser eliminadas através de refinamentos do mesmo. Os clientes

atuam principalmente na especificação do projeto, não participando com ênfase em outras etapas.

Já as metodologias ágeis são centradas nas pessoas participantes do projeto. Elas são definidas por ciclos de desenvolvimento iterativos, os quais são guiados por tomadas de decisões colaborativas, constante *feedback* e integração de mudanças no projeto em desenvolvimento (NERUR; MAHAPATRA; MANGALARAJ, 2005).

O SCRUM é uma metodologia ágil a qual foi adaptada para uso no projeto desenvolvido pelo autor. Ela pode ser visualizada na figura 7 e compreendida em sua totalidade através da explicação dos termos chaves a seguir:

Figura 7 – Fluxo de atividades no *framework* SCRUM



Fonte: scrum.org (SCRUM.ORG, 2022)

1. **Sprint:** período de tempo cíclico e de duração variável de acordo com cada projeto. No projeto desenvolvido pelo autor foram adotados *sprints* de duas semanas;
2. **Sprint planning:** planejamento das atividades a serem realizadas no *sprint*, comumente liderado por gestores de projeto ou *scrum masters*;
3. **Daily scrum:** reuniões diárias de aproximadamente 15 minutos sem uma estrutura definida. Costuma-se relatar progressos nas atividades anteriores e definir as atividades a serem efetuadas no dia vigente;
4. **Sprint review:** reunião de revisão do progressos das atividades estipuladas para o *sprint* vigente, geralmente ocorrendo nos últimos dias do mesmo;

5. **Sprint retrospective:** reunião de análise crítica do *sprint*, na qual se avalia o que funcionou e o que o que necessita de melhoria para o próximo ciclo. Costuma ser o último evento de cada *sprint*;
6. **Product backlog:** fonte que armazena a relação de tarefas que necessitam ser efetuadas para a entrega do produto final;
7. **Sprint backlog:** relação de tarefas a serem realizadas no *sprint* vigente;
8. **Increment:** conjunto de entregas que agregam valor ao produto final. Somatório das funcionalidades adicionadas no *sprint* vigente com aquelas somadas em ciclos anteriores.

A Indicium Tech, empresa na qual o autor realizou o projeto de consultoria para a organização manufatora de baterias, define a duração de seus *sprints* em duas semanas. As *daily scrums* são realizadas no início de cada dia útil, e comumente possuem duração de 15 minutos. Em relação ao modelo *SCRUM* proposto acima foi realizada uma alteração visível: a integração das *sprint reviews* e *sprint restropectives* em uma reunião apenas, a qual acontece na última semana de cada ciclo. Logo, o encontro já concatena uma visão mais operacional do que foi feito e também uma abordagem mais crítica sobre o que pode ser melhorado.

4 ESPECIFICAÇÃO DO PROJETO

No presente capítulo são abordados tópicos referentes ao entendimento detalhado das especificações do projeto. O mesmo está dividido em 3 seções principais: entendimento de negócio, entendimento dos dados e fluxo de modelagem de dados.

4.1 ENTENDIMENTO DE NEGÓCIO

Nesta seção será abordado o fluxo antigo de dados do setor comercial, o escopo geral do projeto, os usuários de negócio que farão uso dos relatórios gerados, os requisitos funcionais e não funcionais da entrega e quais critérios devem ser levados em consideração para determinação do sucesso do trabalho.

4.1.1 Fluxo antigo de dados do setor comercial

As operações de vendas de baterias elétricas são registradas em ERPs (*Enterprise Resource Planning*), os quais são *softwares* de gestão de recursos comumente adotados no mundo empresarial. No projeto descrito nesta monografia, serão consideradas apenas vendas computadas no Brasil, as quais são registradas no ERP Microsoft Dynamics AX. É por intermédio deste *software*, também, que são feitos os cadastros de clientes, produtos, distribuidores, dentre outras entidades relevantes nas operações.

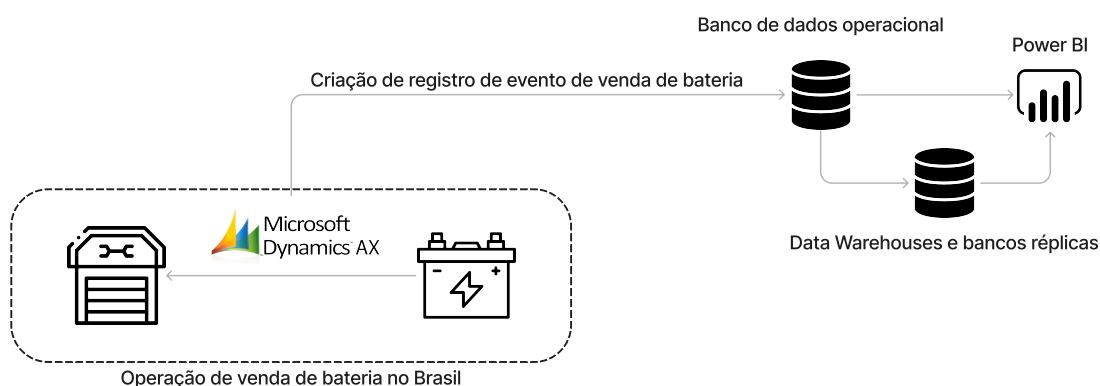
Os dados gerados pelo ERP são armazenados em bancos de dados operacionais, os quais são otimizados para a inserção de registros de eventos em tempo real.

Atualmente, os relatórios comerciais são disponibilizados em um portal hospedado na nuvem. O mesmo atua como um repositório central de relatórios da empresa. Os painéis consomem dados não somente de *data warehouses* distintos, como também do banco de dados operacional. Logo, as transformações de dados acabam sendo feitas no próprio Power BI, concentrando a inteligência de negócio na aplicação. Visualiza-se um resumo do fluxo de dados do setor comercial na figura 8.

4.1.2 Escopo do projeto

A empresa manufatora de baterias elétricas contratante do projeto descrito neste documento, a partir da década passada, começou a investir em projetos de consultorias visando desenvolver maturidade analítica. Foi construído um portal de inteligência que centraliza diversos relatórios da empresa em uma plataforma. Entretanto, cada painel costuma consumir dados de fontes diferentes. Mesmo que alguns relatórios se baseiem em dados estruturados em diferentes *data warehouses*, outros importam dados diretamente de bancos operacionais da empresa, o que resulta em transforma-

Figura 8 – Vendas de baterias registradas no ERP geram dados no banco de dados operacional.



Fonte: autoria própria

ções de dados realizadas diretamente na ferramenta de *business intelligence*, gerando gargalos e quedas de performance nas visualizações.

Logo, definiu-se como meta centralizar as fontes de dados dos principais setores da empresa em um único *data warehouse*. O objetivo principal do projeto é iniciar o processo de centralização das informações da empresa. Com isso, entregar relatórios de *business intelligence* robustos e performáticos. Para este trabalho em questão, foi predefinido que o setor abordado seria o comercial da empresa, com fontes de dados que englobam informações de vendas de baterias das unidades que atendem o Brasil.

4.1.3 Usuários de negócio

Os painéis serão disponibilizados para os seguintes usuários de negócio, contando com limitações nas visualizações de dados através de técnica de governança de dados aplicada no Power BI:

- Diretores globais;
- Gerentes de unidades de distribuidores;
- Vendedores de unidades de distribuidores;
- Determinados clientes.

4.1.4 Requisitos do projeto

Os requisitos de um determinado projeto definem as propriedades e funcionalidades do mesmo. Eles estão divididos em funcionais, os quais descrevem as funciona-

lidades da aplicação, o que ela de fato oferecerá para o usuário final, e não funcionais, os quais que descrevem as restrições e propriedades da implementação.

4.1.4.1 Requisitos funcionais

- Relatórios comerciais: painéis que mostram indicadores comerciais da empresa, como faturamento e volume de baterias vendido;
- Estrutura de dados: dados formatados em um *data warehouse* padronizado;
- Documentação detalhada do projeto: diagramas de relacionamento de dados no *data warehouse* e documentação descritiva de tabelas e colunas na ferramenta dbt;
- Testes unitários de dados: testes diários que alertam quando eventuais informações anômalas são extraídas, como dados duplicados ou nulos.

4.1.4.2 Requisitos não funcionais

- Relatórios devem seguir os seguintes requisitos:
 1. Performance: o carregamento dos visuais nos relatórios deve ser rápido. Da mesma forma não podem haver gargalos na navegação entre as diferentes seções dos painéis;
 2. Atualização diária: os dados devem ser atualizados de forma diária nos painéis de *business intelligence*;
 3. Dados históricos: devem ser adicionados gráficos e tabelas que forneçam ao usuário de negócio uma visão de informações de vendas ao longo do tempo;
 4. Governança de dados: através da técnica *row-level security* do Power BI devem ser limitadas as visualizações de dados de acordo com o endereço de *e-mail* do usuário de negócio;
 5. Interface amigável: as cores que preenchem os botões, gráficos e formas geométricas precisam seguir a identidade visual da empresa contratante. O painel deve ser construído baseado em boas práticas de *storytelling* (sequência em que as informações são dispostas no painel) e UI (interface de usuário);
 6. Navegação facilitada: o manuseio do relatório pelos usuários de negócio deve ser intuitivo, evitando-se que os mesmos necessitem de um conhecimento profundo na ferramenta Power BI para que consigam extrair valor dos painéis.
- Entre as ferramentas utilizadas devem constar:

1. Power BI: a aplicação da Microsoft deve ser utilizada pelo fato de já ser amplamente aceita nas equipes da empresa manufatora de baterias e de permitir customizações de indicadores e visuais;
 2. Microsoft Azure: a computação na nuvem deve ser realizada na plataforma, pois já existem aplicações da empresa sendo executadas na mesma, evitando que elas tenham que ser transferidas e adaptadas para outro sistema. A aplicação engloba funcionalidades de versionamento de código com o Azure Repos e de automação de tarefas com o Azure Pipelines, ambas acessadas via Azure DevOps.
- Arquitetura de dados necessita atender as seguintes premissas:
 1. Tabelas e colunas com nomes humanizados e padronizados: os usuários de negócio precisam entender o conteúdo das tabelas, identificando nas mesmas os processos da empresa;
 2. Atualização diária: os dados armazenados no *data warehouse* devem ser atualizados diariamente com o intuito de que o abastecimento dos painéis de *business intelligence* forneçam indicadores atuais da empresa;
 3. Modelo de dados simplificado em *star-schema*: as tabelas devem respeitar o modelo dimensional, com divisão entre fatos e dimensões, para facilitar a configuração dos relacionamentos de dados no Power BI.

4.1.5 Critérios de sucesso

- Os relatórios devem responder as principais perguntas de negócio dos usuários através de indicadores;
- Dados transformados devem ser validados integralmente para que possam ser consumidos pelos usuários finais;
- Relatórios devem respeitar a identidade visual da empresa contratante do projeto e possuir fácil navegação;
- Deve haver governança de dados nos relatórios com o intuito de limitar o acesso aos dados de acordo com o usuário;
- Novos dados devem ser extraídos diariamente para o *data warehouse*, alimentando uma atualização de mesma frequência no Power BI;
- A documentação do projeto necessita possibilitar que futuras manutenções na transformação dos dados e relatórios possam ser realizadas pelos colaboradores da empresa contratante.

4.2 ENTENDIMENTO DOS DADOS

Após o entendimento do negócio da empresa e seu contexto ter sido concretizado, deve-se realizar o estudo dos dados e o entendimento de como os mesmos retratam os processos da empresa.

Esta seção aborda os materiais que foram fornecidos pela empresa manufatora de baterias para que o autor do presente documento pudesse realizar o entendimento das informações do setor comercial e do relacionamento das tabelas, os indicadores utilizados pelos usuários de negócio para tomadas de decisões e a documentação e diagramas de dados desenvolvidos.

4.2.1 Material disponibilizado

- Consulta na linguagem SQL envolvendo as tabelas do setor comercial e seus respectivos relacionamentos;
- Documento descrevendo os indicadores utilizados pelos usuários de negócio da empresa para tomadas de decisões;
- Acesso ao banco de dados de simulação da empresa contratante para realização de consultas em SQL com o objetivo de compreender as informações que compunham as tabelas;
- Acesso ao portal de inteligência o qual contém os relatórios de *business intelligence* da empresa.

4.2.2 Mapeamento do painel de mapeamento de mercado

Inicialmente, foi fornecido acesso a relatórios consumidos pelos usuários de negócio da empresa do ramo de manufatura de baterias elétricas para os colaboradores da Indicium Tech através do portal de inteligência. O objetivo era possibilitar o mapeamento de como os painéis eram construídos, quais indicadores eram mais relevantes nas análises e encontrar possíveis pontos de melhoria na construção dos mesmos. Para o projeto em questão, o foco foi estabelecido no mapeamento do painel de mapeamento de mercado. A conclusão do estudo do relatório é descrita a seguir:

- Os principais indicadores utilizados no relatório eram o volume de baterias vendido, a meta de volume de vendas e o faturamento;
- Não havia gráficos que ilustrassem a evolução histórica dos indicadores;
- O visual não seguia a identidade visual da empresa contratante;
- Enquanto algumas abas eram poluídas com excessiva quantidade de visuais, outras possuíam espaços não aproveitados.

O mapeamento do painel de mapeamento de mercado feito pelo autor deste documento, aliado ao entendimento dos requisitos de projeto definidos pela empresa manufatora de baterias, precedeu a implementação de um novo relatório.

4.2.3 Indicadores mapeados

- Volume vendido: número de unidades de baterias vendidas;
- Meta de volume: meta de unidades de baterias a serem vendidas;
- Desvio de volume: volume vendido - meta de volume;
- Faturamento: valor financeiro em reais referente às vendas de baterias.

4.2.4 Criação de documentação e modelos de dados

Foram descritas, em planilhas, todas as tabelas fontes geradas a partir do ERP Dynamics AX e suas respectivas colunas. Também foram mapeadas informações como contagem de linhas por tabela.

Em seguida, após o entendimento dos relacionamentos entre as tabelas fontes, decidiu-se quais iriam compor as tabelas de dimensões e de fatos.

- Tabelas de dimensões:
 - CFOPs: mapeia os Códigos Fiscais de Operações e Prestações utilizados pela empresa. As CFOPs identificam a operação de cada nota fiscal. Cada linha desta tabela de dimensões corresponde a uma operação, sua descrição e se as notas fiscais que a possuem devem ser levadas em consideração no volume e faturamento na tabela de fatos de faturas;
 - Clientes: identifica os clientes da empresa e respectivas informações dos mesmos, como segmentos de venda, endereços e seus nomes completos;
 - Datas: cria um intervalo de datas o qual abrange todo o período englobado pelas tabelas de fatos. Contém informações complementares de cada data, como o trimestre, dia da semana e se a mesma é dia útil;
 - Famílias comerciais: contém as famílias comerciais da empresa manufatora de baterias elétricas, seus códigos e quais os tipos de baterias elas abrangem. Elas definem se os produtos pertencem à categoria automotiva, estacionárias, dentre outras;
 - Unidades, equipes e áreas: identifica as unidades de distribuidores da empresa no Brasil, as quais contêm equipes de vendedores. Cada time é dividido em áreas, nas quais estão localizados os vendedores. Enquanto os supervisores de cada área podem visualizar informações das vendas do

seu grupo, eles não devem ter acesso à indicadores de outras áreas. Logo, deve ser aplicada uma limitação na visualização dos indicadores através do *row-level security* no Power BI.

- Tabelas de fatos:
 - Faturas: cada linha desta tabela possui informações do volume negociado em uma operação e o valor referente ao mesmo. Contém chaves estrangeiras de CFOPs, clientes, datas, famílias comerciais, produtos e unidades, equipes e áreas. Como descrito na seção 3.5.1, tais chaves estrangeiras se comunicam com as primárias das tabelas de dimensões;
 - Metas de volumes: contém informações de metas de volume de vendas das unidades da organização. Cada linha possui informações do valor da meta de volume a ser alcançada, contendo chaves estrangeiras de datas, famílias comerciais e unidades, equipes e áreas.

Como determinado na seção 3.5.1, os atributos das tabelas de dimensões funcionam como restrições ao cálculo dos indicadores nas tabelas de fatos. Logo, consegue-se visualizar o volume vendido, faturamento e meta de volume por cada coluna das tabelas de dimensões, desde que haja um relacionamento entre as tabelas. Tratando-se de visualizações no Power BI, os indicadores dos modelos fatos irão compor o eixo das ordenadas, enquanto os atributos das dimensões o eixo das abcissas.

Por fim, foram elaborados diagramas baseados na metodologia *star-schema* para documentar a arquitetura de dados e facilitar a futura transformação dos mesmos seguindo requisitos da modelagem dimensional, como se pode observar na figura 9.

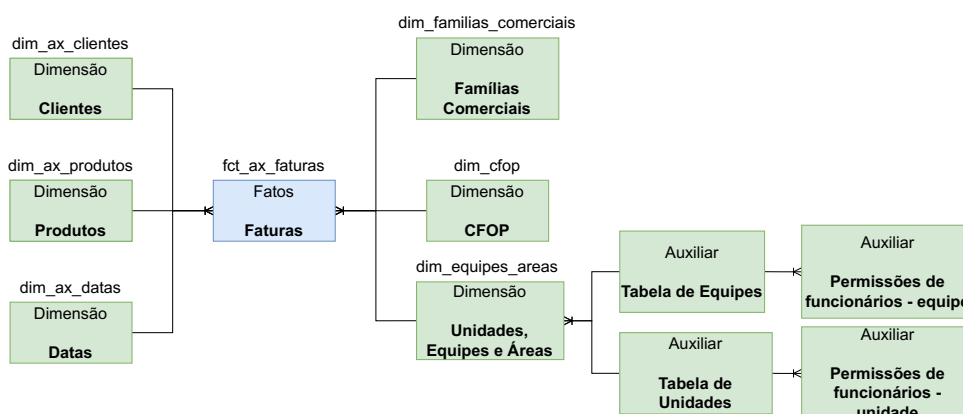
4.3 FLUXO DE MODELAGEM DE DADOS

A definição da sequência de transformações a serem realizadas nos dados não somente auxilia o desenvolvedor das mesmas a compreender melhor os próximos passos de desenvolvimento, mas também resulta em uma arquitetura de dados mais organizada e modular.

A divisão da transformação de dados em diferentes etapas é facilitada pela modularidade da ferramenta dbt. A mesma permite que as distintas tabelas criadas a partir de consultas no formato SQL sejam referenciadas em outros arquivos. Pode-se observar na figura 10 o sequenciamento da transformação de dados na aplicação.

Nas subseções a seguir serão abordadas as divisões das áreas de transformação de dados de acordo com o tipo de tratamento que as mesmas abordam, o versionamento de código seguindo conceitos do modelo *git flow* e a separação de ambientes de desenvolvimento, testes e produção.

(a) Diagrama de implementação do *star-schema* de faturas.



(b) Diagrama de implementação do *star-schema* de metas de volumes.

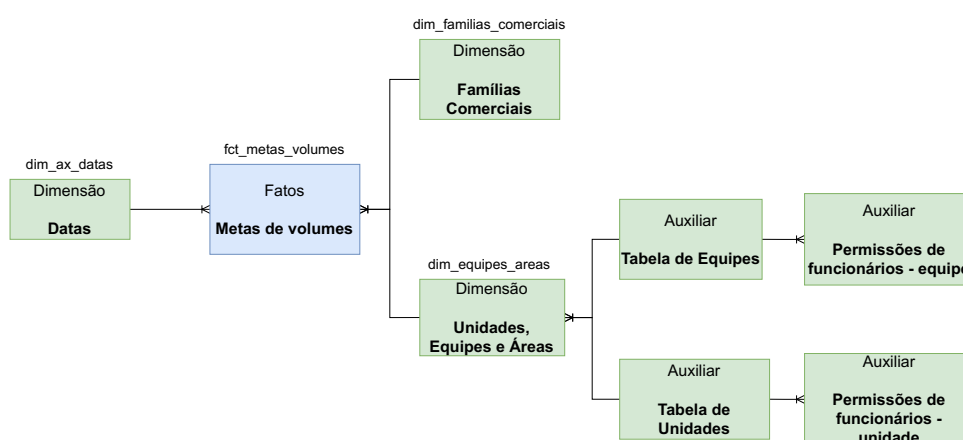


Figura 9 – Diagramas de implementação do *star-schema* dos dados do ERP AX.

Fonte: autoria própria

4.3.1 Área *staging*

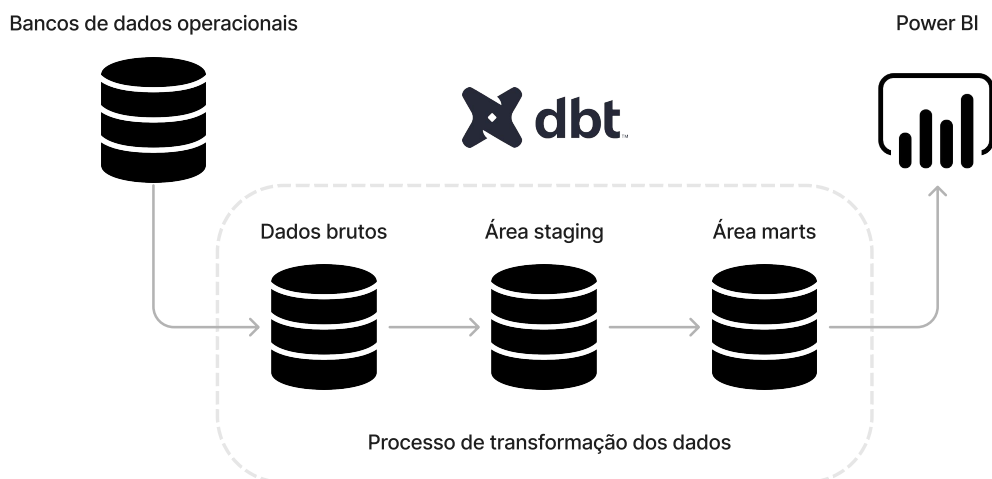
As tabelas *staging* são criadas a partir das tabelas fontes extraídas dos bancos operacionais do cliente. Uma boa prática é referenciar apenas uma tabela fonte em cada tabela *staging* com o objetivo de limitar a transformação nesta etapa a tratamentos básicos, sem mudanças na estrutura global da tabela.

Os diferentes tipos de transformação que devem ser aplicados nas tabelas da área *staging* do projeto descrito nesta monografia estão descritos a seguir.

4.3.1.1 Renomeação de colunas

Humanização dos nomes de coluna e tradução das nomenclaturas para a língua portuguesa para fácil entendimento dos usuários de negócio.

Figura 10 – Fluxo dos dados no dbt.



Fonte: autoria própria

4.3.1.2 Mudança de tipos de dados

Tipo de dado especifica o tipo de valor que uma variável armazena e quais operações podem ser aplicadas na mesma (DATA. . . , 2016). No projeto descrito neste documento, pode-se aplicar transformações de tipos de dados textuais para numéricos e datas, por exemplo.

4.3.1.3 Tratamento de dados

Substituição de valores vazios para nulos com o objetivo de ocupar menos espaço no banco e evitar falhas em ligações de tabelas, transformação de dados textuais escritos em letra minúscula para maiúscula com o intuito de padronizar a visualização desses elementos e eliminação de eventuais registros duplicados.

4.3.1.4 Seleção de colunas

Seleção de colunas que poderão ser aproveitadas nas etapas seguintes da transformação de dados e que irão compor os relatórios. Tal fator contribui para a otimização no tempo de geração e de consultas dessas tabelas e para tornar sua estrutura mais simplificada, fato que contribui para o melhor entendimento das mesmas por parte dos usuários de negócios.

A área *staging* tem grande importância no projeto ao contribuir para a modularidade do mesmo, pois ela concentra a maior parte dos tratamentos básicos de dados os quais serão referenciados na etapa seguinte do fluxo de transformação dos dados.

4.3.2 Área *marts*

As tabelas *marts* são construídas a partir de tabelas *staging* cujos dados já receberam tratamentos iniciais. Nesta etapa ocorrem agregações de dados, ligação de tabelas distintas, criação de colunas calculadas e chaves primárias e estrangeiras compostas (quando a ligação entre tabelas distintas ocorre por intermédio de mais de uma coluna).

Os *scripts* de transformação nesta área são responsáveis pela geração das tabelas que compõem o modelo dimensional (fatos e dimensões). Logo, é a última etapa de tratamento dos dados dentro da ferramenta dbt.

A estruturação do repositório do dbt em *staging* e *marts* é amplamente aplicada em projetos de transformação de dados e recomendada pelos próprios funcionários da empresa desenvolvedora da aplicação, como descrito em (CARROLL, 2019).

4.3.3 Versionamento de código

Com o intuito de viabilizar o trabalho de diversos colaboradores em um mesmo repositório do dbt e a divisão de ambientes de desenvolvimento, testes e produção, utilizou-se o sistema de controle de versão git. Git é utilizado para rastreamento de mudanças em arquivos, comumente aplicado no gerenciamento de códigos em desenvolvimento de *softwares* (PERVEEZ, 2022).

No trabalho descrito neste documento, o repositório do dbt foi versionado com o sistema git através de funcionalidade da aplicação Azure DevOps, o Azure Repos. Além de permitir que o usuário se beneficie do versionamento de arquivos através do git, a ferramenta também disponibiliza outras funcionalidades como o Azure Pipelines, o qual também foi utilizado no projeto e que permite a orquestração e automação de tarefas.

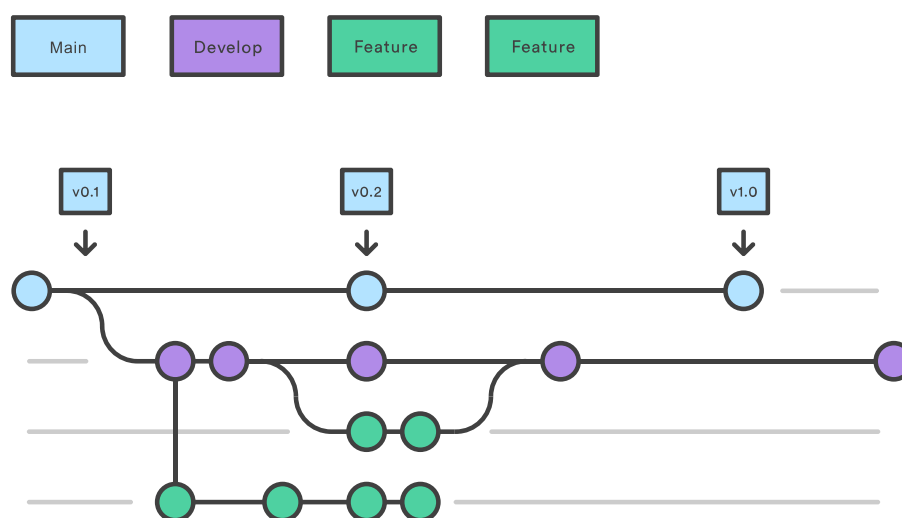
Por fim, foi utilizado o modelo de fluxo de trabalho gitflow para segmentação dos ambientes de desenvolvimento, teste e produção. Ele consiste na criação de três ramificações principais:

- *Main: branch* que armazena a versão do repositório em produção, e que só deve ser atualizado após a aprovação de *pull request* e os modelos terem sido testados na *branch develop*;
- *Develop*: ramificação intermediária de desenvolvimento a qual armazena o repositório com as alterações realizadas pelos usuários nas *branches feature* e que também apenas deve ser atualizada via *pull request*;

- *Feature*: conjunto de *branches* criadas a partir da *develop* com o objetivo de viabilizar que o usuário desenvolva novas funcionalidades sem afetar as ramificações de desenvolvimento e produção citadas.

Deve-se salientar que, enquanto as ramificações *develop* e *main* são utilizadas durante todo o desenvolvimento do projeto e recebem constantes atualizações de arquivos através de *pull requests*, cada *branch feature* é criada para desenvolvimento de determinada funcionalidade. O fluxo pode ser visualizado na figura 11.

Figura 11 – Fluxo de trabalho gitflow.

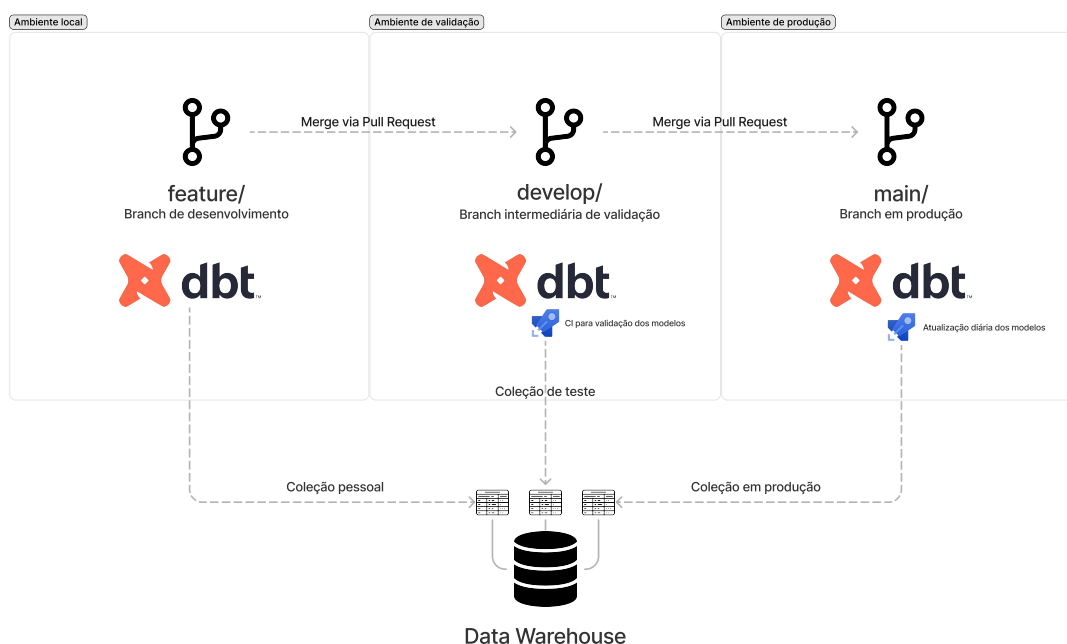


Fonte: (ATLASSIAN.COM, 2022)

4.3.4 Ambientes de desenvolvimento, validação e produção

Com o intuito de que a transformação de dados em fase de implementação não afete as tabelas em produção (disponíveis para o usuário final), foram criados três ambientes distintos, os quais possuem *data schemas* (coleções de dados) individuais no *data warehouse* e geram tabelas com base nos repositórios de cada ramificação do *gitflow*. O fluxo de desenvolvimento nos ambientes está ilustrado no diagrama da figura 12.

Figura 12 – Funcionamento dos diferentes ambientes do projeto e integração com as ramificações do gitflow.



Fonte: autoria própria

4.3.4.1 Ambiente de desenvolvimento

O desenvolvimento das transformações de dados é realizado em máquinas virtuais disponibilizadas pela equipe da empresa contratante com o intuito de evitar que os dados da empresa trafeguem pela rede de *internet*.

Durante o desenvolvimento das transformações de dados, testes e documentação do projeto, o repositório do dbt é versionado em ramificações independentes do tipo *feature*. Logo, publicações de alterações no repositório Azure Repos na nuvem não afetam as *branches* de validação e produção.

Para fins de testes, é recomendado que os colaboradores gerem as tabelas em desenvolvimento para eventuais validações de dados e para que testes sejam executados referenciando as mesmas. Elas são materializadas em coleções de dados individuais no *data warehouse* para cada desenvolvedor das transformações, funcionalidade a qual é configurada no próprio dbt através da definição de perfis de usuários.

Após as modificações e implementações de código terem sido finalizadas e validadas no repositório local do dbt através de uma ramificação *feature* local, deve-se fazer a publicação da mesma na nuvem. Dessa forma, é possível abrir um *pull request* das funcionalidades para que, após o mesmo ser aprovado por outros colaboradores do projeto, ele seja concluído e as alterações e implementações no repositório sejam

aplicadas na ramificação de desenvolvimento.

4.3.4.2 Ambiente de validação

O ambiente de validação é composto pela ramificação *develop*, a qual recebe as implementações e modificações realizadas no ambiente de desenvolvimento com o objetivo de testá-las antes que as mesmas sejam consumidas no ambiente final de produção. Tais testes são criados no dbt e executados na nuvem através da ferramenta *Azure Pipelines*. Após resultado positivo dos testes, é possível criar um novo *pull request* para a ramificação *main*, a qual contém os códigos em SQL finalizados e validados os quais irão gerar as tabelas consumidas pelo Power BI. Caso haja falhas nos testes, deve-se criar uma nova *branch* no ambiente de teste para realizar eventuais alterações.

4.3.4.3 Ambiente de produção

O ambiente de produção é aquele que contém a ramificação do repositório do dbt previamente validado e testado. Os *scripts* SQL em produção geram as tabelas em uma coleção de dados disponibilizada para o cliente, a qual irá abastecer os relatórios na aplicação Power BI.

Como um dos requisitos do cliente é a atualização diária dos dados, deve ser definida a automação da geração de tabelas na ferramenta *Azure Pipelines*.

Deve-se ressaltar que, embora o ambiente de produção tenha sido previamente testado e validado com o auxílio das funcionalidades do dbt e *Azure Pipelines*, eventuais erros podem acontecer na atualização diária dos dados. Logo, é papel do autor deste documento investigar, com o auxílio de um engenheiro de dados, os motivos das falhas e realizar manutenções nos códigos no dbt para correção.

Coube ao autor desta monografia no processo de criação dos ambientes a configuração dos perfis no dbt e configuração do repositório no Azure Repos. O mesmo utilizou os ambientes para desenvolvimento e publicação das transformações de dados, testes e documentação do projeto.

5 IMPLEMENTAÇÃO DO PROJETO

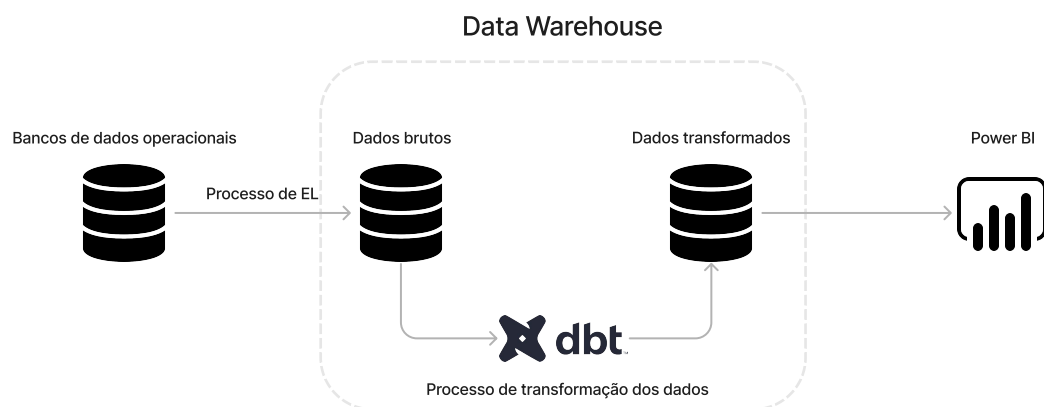
Este capítulo descreve a efetiva implementação do projeto pelo autor, o qual aplicou conceitos e metodologias abordadas no capítulo 3, respeitando as especificações estabelecidas anteriormente.

O presente capítulo está dividido em 3 seções principais: transformação de dados, criação de testes de dados e documentação e desenvolvimento do painel de *business intelligence*.

5.1 TRANSFORMAÇÃO DOS DADOS

Após o entendimento dos dados, definição de quais irão compor as tabelas de fatos ou dimensões e, por fim, a criação dos diagramas de arquitetura dimensional seguindo a metodologia *star-schema*, foi possível iniciar o processo de transformação dos dados seguindo o fluxo descrito na seção 4.3.

A primeira etapa foi a configuração das dependências da ferramenta dbt, a qual hospedou todo o processo de transformação dos dados e possibilitou que o mesmo fosse implementado em camadas através das áreas *staging* e *marts*. Pode-se observar na figura 13 o curso dos dados desde a extração dos mesmos a partir de bancos de dados operacionais até a sua utilização na aplicação Power BI.

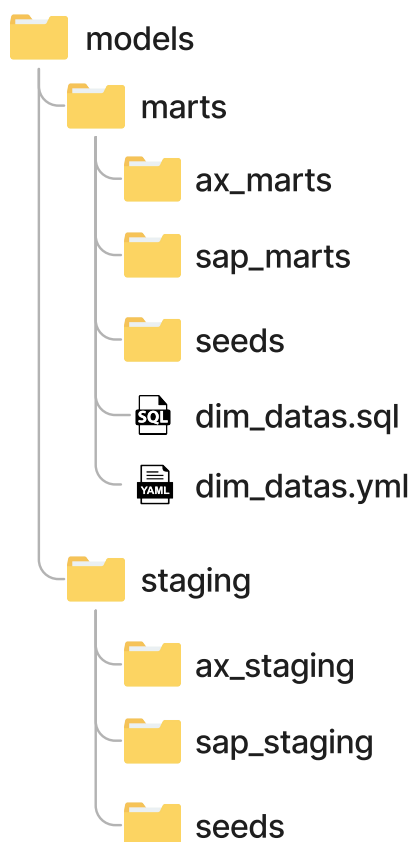


Fonte: autoria própria

5.1.1 Organização do repositório do dbt

Como previamente descrito na seção 3.6, os arquivos de consultas de bancos de dados responsáveis pela transformação dos dados são armazenados na pasta *models* no dbt, a qual é segmentada em dois diretórios: *staging* e *marts*. Os mesmos são divididos em pastas de acordo com a fonte de dados que seus respectivos arquivos referenciam. É possível visualizar na figura 14 como o autor estruturou a pasta contendo os modelos de tratamento de dados. Diretórios com prefixos *ax_* possuem transformações de dados provenientes do banco de dados operacional do ERP Microsoft Dynamics AX utilizado pela empresa contratante enquanto aqueles com o prefixo *sap_* possuem *scripts* de tratamentos de dados provenientes do banco de dados operacional do SAP ERP, e não entram no escopo do projeto em questão. Por fim, os diretórios nomeados como *seeds* são compostos por arquivos os quais transformam dados provenientes de planilhas no formato CSV.

Figura 14 – Organização da pasta *models* do dbt.



Fonte: autoria própria

Na figura 15 é possível verificar como arquivos SQL e YML compõem os sub-diretórios respectivos à cada fonte de dados distinta. Foi considerada apenas uma amostra de arquivos para um entendimento facilitado por parte do leitor.

A área *staging*, a qual atua como camada inicial de transformação de dados, armazena o arquivo *sources.yml*, este que, além de configurar as tabelas dos bancos de dados fontes, fornece descrição das mesmas e realiza testes com o intuito de verificar se a extração diária de novos dados tem funcionado. Já os arquivos YML nas pastas *marts* são criados com o objetivo de documentar as tabelas de respectivos nomes e suas colunas e definir testes de dados em colunas chaves.

Figura 15 – Amostra de arquivos que compõem as pastas *staging* e *marts* no dbt.



Fonte: autoria própria

Por fim, verifica-se na figura 16 como acontece a integração das tabelas de diferentes etapas no dbt.

5.2 CRIAÇÃO DE TESTES DE DADOS E DOCUMENTAÇÃO

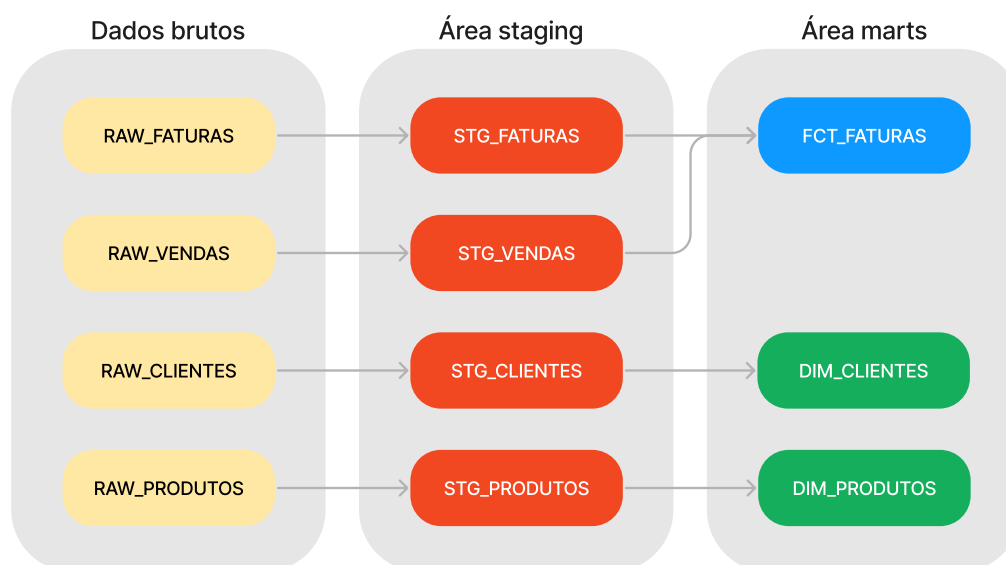
Enquanto os testes de dados têm papel de alertar a equipe de colaboradores quando eventuais anomalias são detectadas nas tabelas, a documentação do projeto é desenvolvida com o intuito de facilitar o entendimento dos dados para novos usuários de negócio e colaboradores e de ser a fonte de respostas de eventuais dúvidas quanto à transformação dos dados.

O desenvolvimento de testes de dados e documentação, o qual é feito nos arquivos YML armazenados no repositório do dbt, será descrito nas subseções seguintes.

5.2.1 Testes de dados

No projeto descrito nesta monografia, testes de dados foram desenvolvidos para verificar diariamente se as extrações de novos dados ocorreram com sucesso e se as

Figura 16 – Etapas de transformação das tabelas do modelo dimensional no dbt.



Fonte: autoria própria

chaves primárias e estrangeiras das tabelas de fatos e dimensões correspondem a determinados requisitos.

- Testes de verificação de extrações diárias: implementados na configuração das tabelas fontes brutas com o objetivo de verificar se a extração diária ocorreu como previsto. Caso a defasagem nos dados seja apenas de 1 dia, uma mensagem de alerta é gerada pelo dbt. Entretanto, quando o atraso é maior que tal período, uma mensagem de erro é emitida;
- Testes de unicidade: aplicados em chaves primárias de tabelas, verificam se as mesmas, as quais podem ser baseadas na combinação de duas ou mais colunas, contêm apenas registros únicos. Caso valores duplicados sejam encontrados, alertas de erros são gerados para o usuário;
- Testes de não-nulidade: também aplicados em chaves primárias, identificam quando há valores nulos nas mesmas;
- Testes de relacionamento: implementados em chaves estrangeiras de tabelas de fatos, verificam se o registro pode ser encontrado na chave primária de uma tabela de dimensão especificada no próprio teste. Verifica se as tabelas estão atendendo aos requisitos de modelagem conceitual.

Coube ao autor desta monografia implementar os testes para as transformações de dados oriundos do ERP Microsoft Dynamics AX através dos arquivos YML do dbt.

5.2.2 Documentação da transformação de dados

A documentação de projetos de *softwares* pode ser considerada como um meio de comunicação entre os colaboradores envolvidos no desenvolvimento e os usuários das aplicações (CHOMAL; SAINI, 2015). Entretanto, tal conceito não se aplica somente a projetos de *software*, como também a projetos de *analytics* como o descrito neste documento.

No projeto em questão, a elaboração da documentação teve como objetivo, além de permitir que colaboradores novos no projeto compreendessem os dados e as motivações por trás das transformações dos mesmos, a transferência de conhecimentos do projeto possuídos pelos colaboradores para um local central onde os usuários de negócio pudessem consultar sempre que necessário.

A documentação do trabalho em questão foi gerada a partir da funcionalidade dbt Docs.

5.2.2.1 dbt Docs

O dbt Docs é uma página *web* gerada por comandos no próprio ambiente do dbt. A funcionalidade hospeda todas as descrições de tabelas, colunas e do projeto no geral desenvolvidas em arquivos de extensão YML e MD (*markdown*).

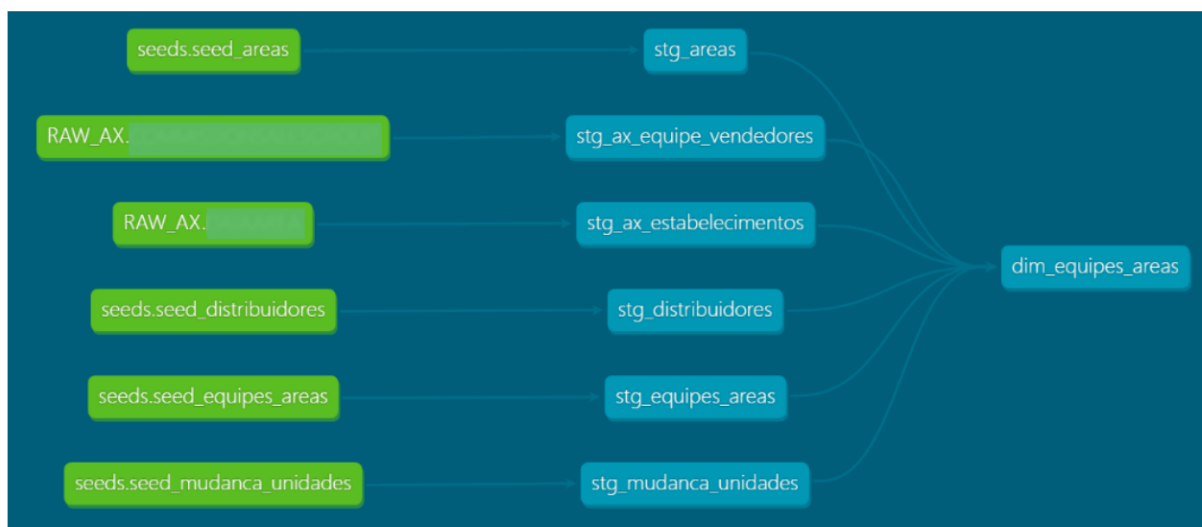
Em arquivos de extensão YML os quais recebem nomes idênticos aos modelos em SQL, há a documentação do significado de negócio da tabela e de suas colunas. Da mesma forma, são descritas eventuais métricas e colunas calculadas. Já os arquivos *markdown* documentam o projeto, seus objetivos e boas práticas no geral.

Por fim, a plataforma dbt Docs também gera automaticamente um diagrama que ilustra como as tabelas de diferentes camadas de transformação de dados se comunicam entre si. Tal funcionalidade é muito útil para entender o impacto de que eventuais modificações em um modelo podem causar em outros. Um exemplo de amostra do projeto pode ser visto na figura 17.

5.3 DESENVOLVIMENTO DO PAINEL DE *BUSINESS INTELLIGENCE*

A elaboração do relatório de mapeamento de mercado na aplicação Power BI foi o último passo de implementação do projeto como um todo. O processo de desenvolvimento pode ser dividido em mapeamento de painéis já consumidos por colaboradores da organização manufatora de baterias, criação de esboço do painel, configuração das dependências do Power BI, desenvolvimento da interface e sua

Figura 17 – Amostra de diagrama gerado automaticamente na ferramenta dbt Docs.



Fonte: autoria própria

navegabilidade e a efetiva implementação dos visuais de indicadores, tópicos os quais serão abordados nas subseções a seguir.

5.3.1 Criação de esboço do painel

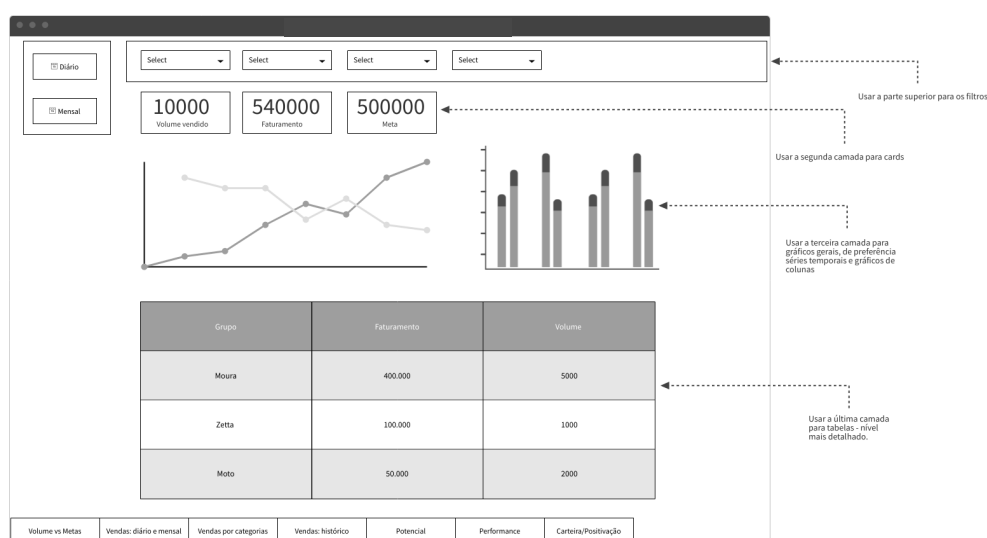
Criou-se um esboço do painel com o objetivo de definir como as diferentes informações seriam distribuídas dentro de cada aba do relatório de mapeamento de mercado. O mesmo teve relevância no sentido de definir como a narrativa dos dados deveria ser inserida no painel. A seguir, são descritos os tipos de visuais na ordem recomendada de apresentação dos mesmos.

1. Filtros: inseriu-se os botões de filtros (nomeados como segmentações de dados no Power BI) na parte superior do esboço do relatório com o intuito de fornecer fácil acesso aos usuários de negócio;
2. Cartões: os *cards* mostram valores de colunas e métricas calculadas de forma literal. No esboço do painel, os mesmos foram inseridos na parte superior com a função de fornecer ao usuário de negócio uma visão global de volume, meta de volume e faturamento da empresa;
3. Gráficos de linhas: adotou-se os gráficos de linhas para plotar indicadores em séries temporais. Inseridos abaixo dos cartões, os mesmos têm função informar como as métricas se comportaram no tempo;
4. Gráficos de barras e colunas: inseriu-se os gráficos de barras e colunas com o intuito de plotar o comportamento dos indicadores segmentado pelos atributos das dimensões;

5. Tabelas: por fim, escolheu-se adicionar tabelas na parte inferior do relatórios com o intuito de as mesmas fornecerem uma visão mais detalhada dos indicadores.

Visualiza-se, na figura 18, o esboço que foi elaborado pelo autor no projeto em questão. Deve-se notar como os dados são apresentados de forma mais global na parte superior, contrastando com visualizações detalhadas na seção inferior.

Figura 18 – Esboço desenvolvido para proposta de narrativa de dados.



Fonte: autoria própria

5.3.2 Configuração do Power BI

Na ferramenta Power BI, o primeiro passo é a configuração das fontes de dados e seus relacionamentos. Após a conexão com o *data warehouse* modelado ter sido feita, é possível selecionar quais tabelas devem ser importadas. Escolheu-se apenas tabelas cujos dados teriam relevância no desenvolvimento do relatório de mapeamento de mercado com o objetivo de não poluir o modelo de dados da ferramenta e de entregar um painel performático.

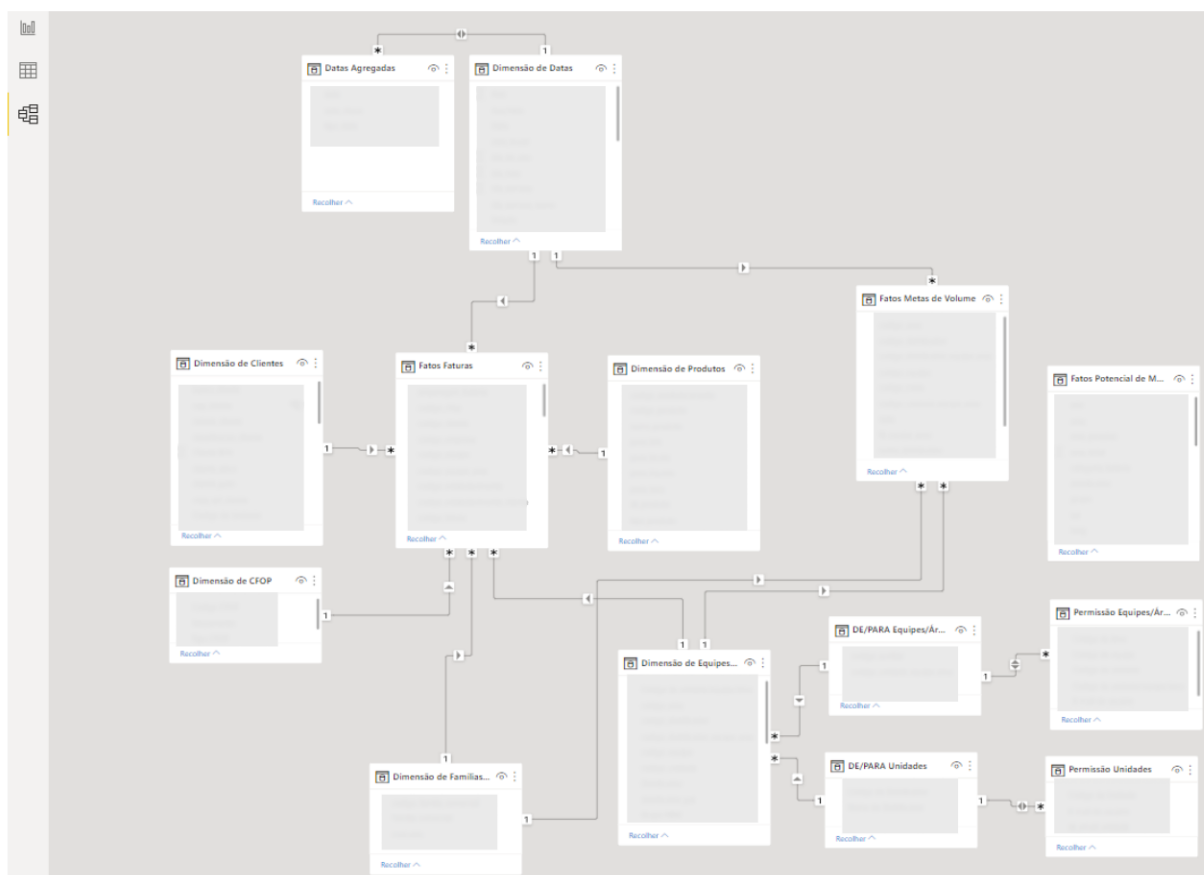
5.3.2.1 Modelo de dados

A configuração do modelo de dados consiste em criar relacionamentos entre as tabelas de fatos e dimensões transformadas. A conexão entre as tabelas deve ser

feita através da ligação de chaves estrangeiras das tabelas de fatos com as chaves primárias das tabelas de dimensões. Tais relacionamentos são ditos como muitos para um (*:1), o que significa que, enquanto as chaves estrangeiras das tabelas de fatos possuem valores repetidos, as chaves primárias das tabelas de dimensões têm dados únicos.

Visualiza-se, na figura 19, como foi realizada a conexão entre as diferentes tabelas do modelo. Nota-se, também, a similaridade com os diagramas *star-schemas* apresentados na figura 5.

Figura 19 – Configuração das fontes de dados no Power BI.



Fonte: autoria própria

5.3.2.2 Conexão de dados

É possível, no Power BI, consumir os dados de distintas formas:

- *DirectQuery*: consultas ao bancos de dados fonte são feitas assim que há a criação de visuais e interação com os mesmos. Logo, os dados não ficam armazenados no próprio Power BI;

- Importação de dados: as tabelas são importadas e armazenadas no Power BI. Para constante atualização das informações do relatório, este método depende de atualizações diárias na importação dos dados;
- Misto de *DirectQuery* e importação de dados: serviço destinado a usuários *premium* da ferramenta, e que possibilita a integração dos dois métodos citados acima. Assim, é possível armazenar dados históricos e realizar consultas apenas para obter informações recentes, obtendo-se ganho na performance.

Inicialmente, no projeto em questão, os dados eram consumidos através do método *DirectQuery*. Entretanto, por se tratar de consultas de tabelas populosas, começaram a surgir gargalos críticos na performance do carregamento de visuais no painel. Logo, optou-se pela troca do consumo dos dados para o método de importação, o que possibilitou que os visuais sejam carregados em um intervalo de tempo expressivamente menor.

5.3.3 Desenvolvimento da interface de navegação

Optou-se pela divisão do relatório em diferentes abas para evitar a poluição do relatório e para que o mesmo não se tornasse extenso.

Com o intuito de facilitar a navegação entre as diferentes abas do painel, o autor implementou botões na parte superior do relatório. Também foi inserido um botão na seção superior direita que fornece acesso aos filtros de segmentação dos dados.

Tais implementações visam fornecer ao usuário de negócio uma navegabilidade fluida entre as diferentes seções do relatório.

5.3.4 Implementação dos visuais

Neste tópico, são abordadas em detalhes as diferentes seções do painel de mapeamento de mercado. Com o intuito de preservar os dados da empresa contratante, multiplicou-se os indicadores presentes nas figuras dos tópicos a seguir por constantes ínfimas e aleatórias. Da mesma forma, ocultou-se o logo da empresa nas diversas abas do relatório, além de nomes de clientes e produtos da mesma.

5.3.4.1 Aba geral

Desenvolveu-se a aba geral com o intuito de oferecer ao usuário uma visão global dos indicadores da empresa. Ela engloba o acompanhamento dos valores dos KPIs através de cartões, da evolução histórica dos indicadores por intermédio de um gráfico de linhas, além de um resumo das métricas agregadas por dimensões importantes, como clientes e famílias comerciais (as quais se referem ao mercado

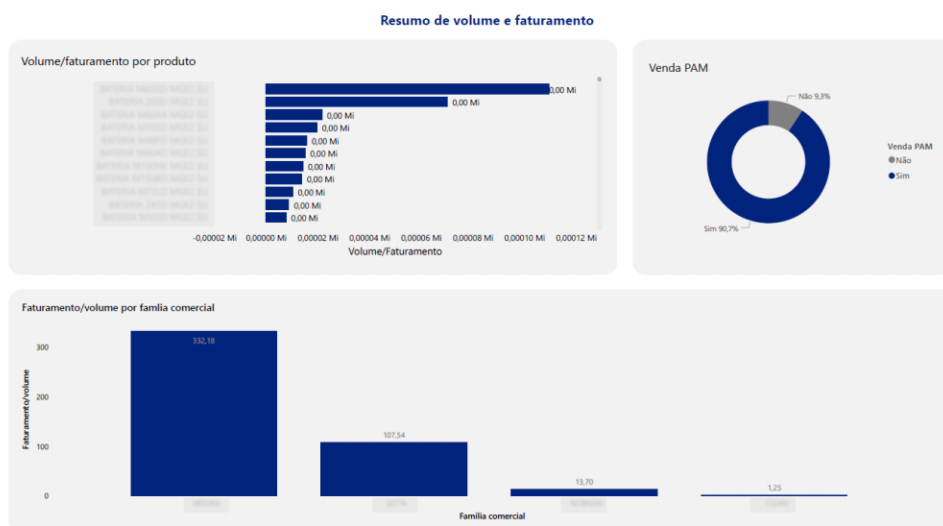
que determinada bateria está inserida). A interface da seção geral pode ser vista nas figuras 20 e 21.

Figura 20 – Parte 1 da aba geral do painel Mapeamento de Mercado.



Fonte: autoria própria

Figura 21 – Parte 2 da aba geral do painel Mapeamento de Mercado.



Fonte: autoria própria

5.3.4.2 Aba de acompanhamento da meta de volume

A seção de acompanhamento de meta de volume foca na comparação entre os indicadores de volume de baterias vendido e de meta. Além de plotar a evolução dos

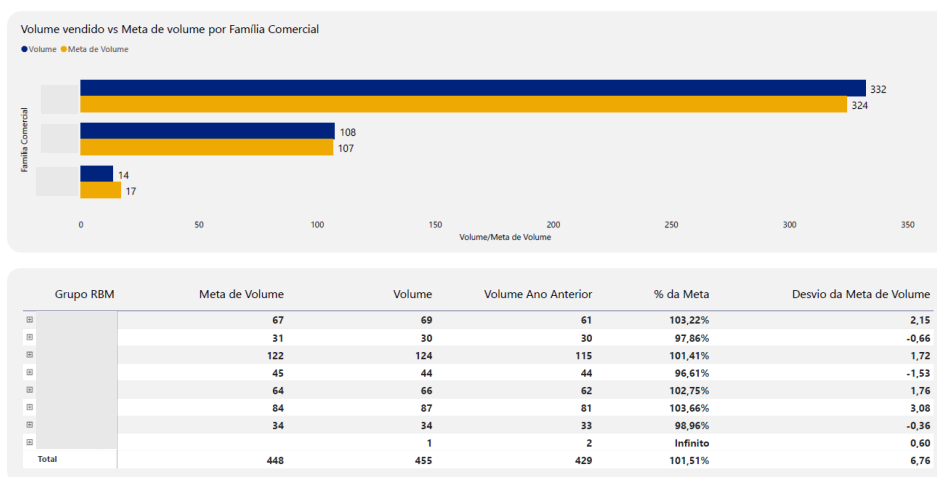
mesmos em intervalos no formato ano/mês, mostra a comparação dos KPIs para cada família comercial. Por fim, em um visual no formato de tabela, mostra a comparação dos indicadores para distintas divisões internas de grupo de vendedores da empresa. Visualiza-se sua interface na figura 22 e 23.

Figura 22 – Parte 1 da aba de acompanhamento de meta de volume do painel Mapeamento de Mercado.



Fonte: autoria própria

Figura 23 – Parte 2 da aba de acompanhamento de meta de volume do painel Mapeamento de Mercado.

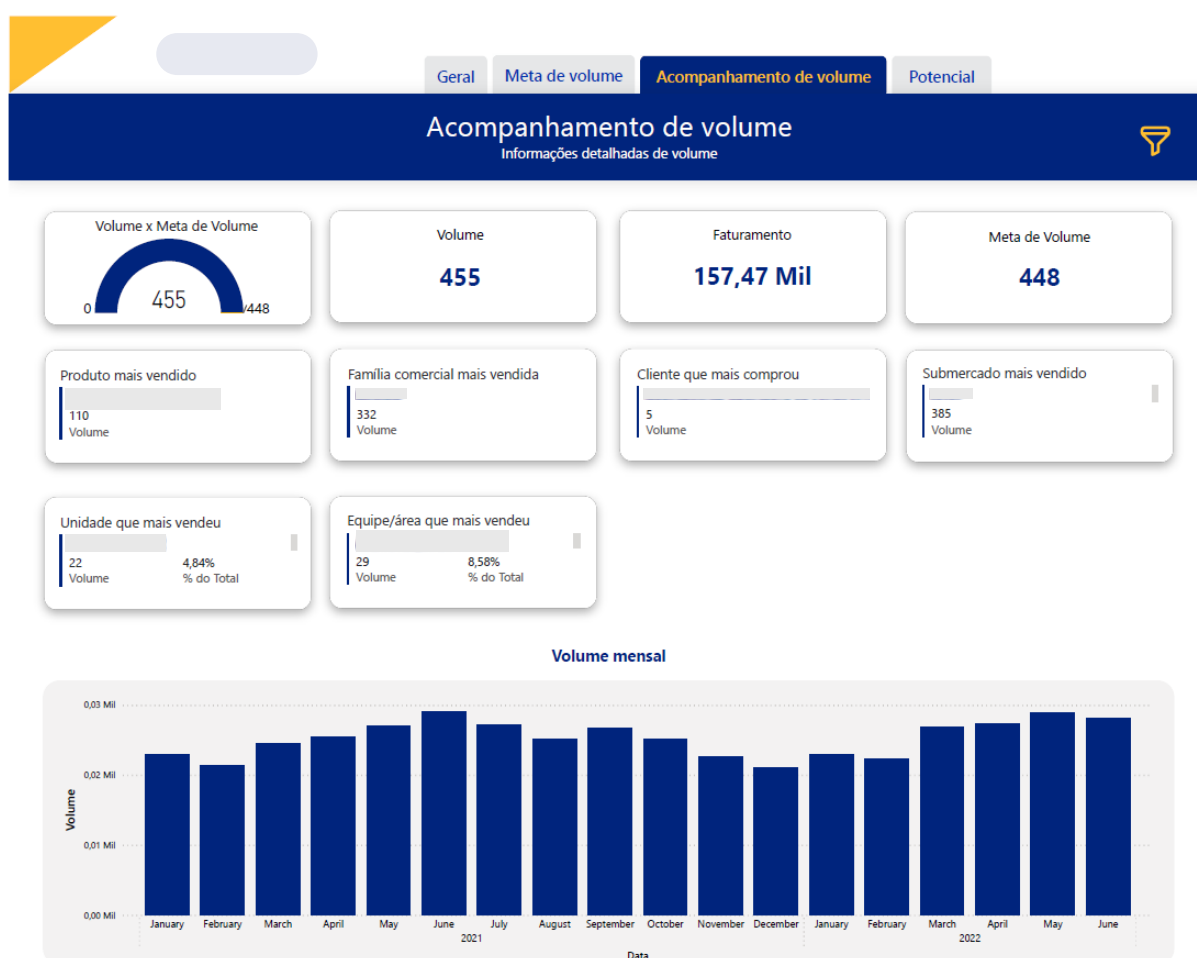


Fonte: autoria própria

5.3.4.3 Aba de acompanhamento detalhado de volume

Visualiza-se, na aba de acompanhamento detalhado de volume, como o indicador de volume de vendas de baterias agregado por distintos atributos de dimensões se comporta. É apresentado o volume de vendas por produto, cliente, família comercial e distribuidores da empresa, além de comparações históricas dos dados no formato YoY (*year over year*).

Figura 24 – Parte 1 da aba de acompanhamento detalhado de volume do painel Mapeamento de Mercado.

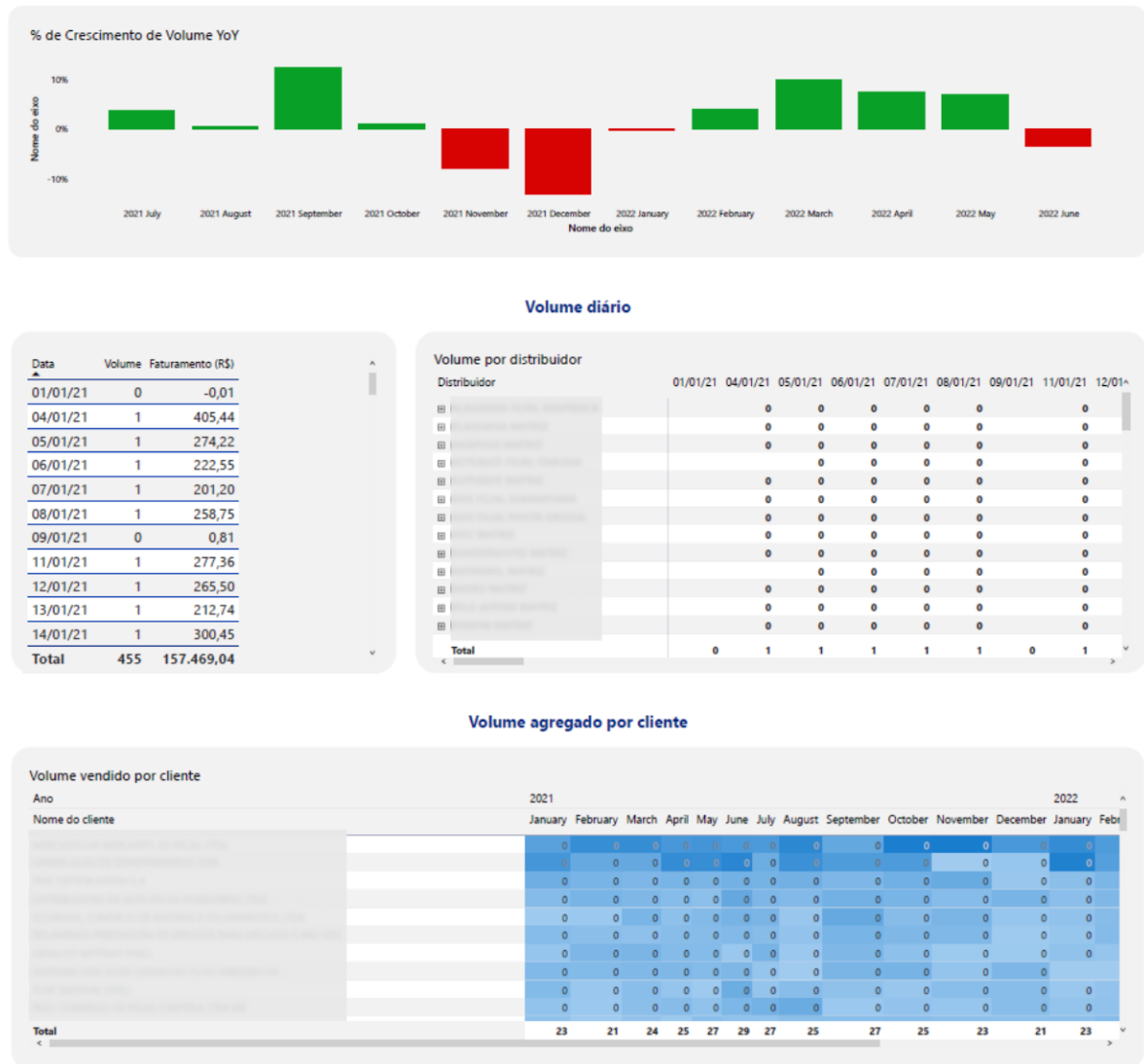


Fonte: autoria própria

5.3.4.4 Seção de filtros

A seção de filtros de segmentação de dados é acessível a partir de qualquer aba do relatório pelo botão localizado na parte superior direita. Ela possibilita que o usuário restrinja as visualizações dos relatórios a partir de atributos de tabelas de dimensões, como datas, cidades, famílias comerciais, segmentos de vendas, dentre outros. Pode-se observar na figura 27 a disposição dos filtros.

Figura 25 – Parte 2 da aba de acompanhamento detalhado de volume do painel Mapeamento de Mercado.

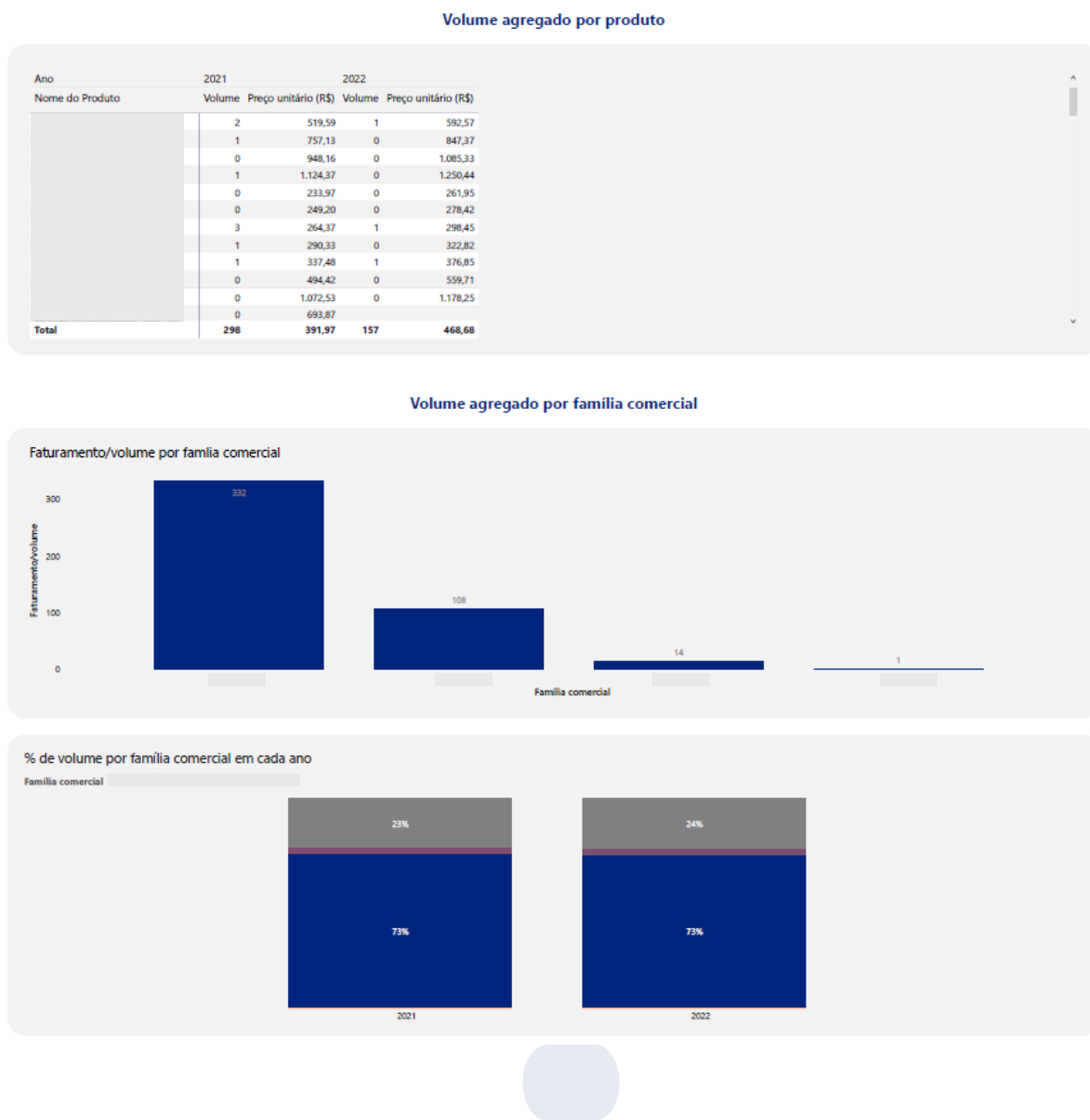


Fonte: autoria própria

A inserção de filtros em uma camada oculta acessível por um botão torna o relatório mais organizado e viabiliza que o usuário foque sua atenção na visualização dos indicadores em si.

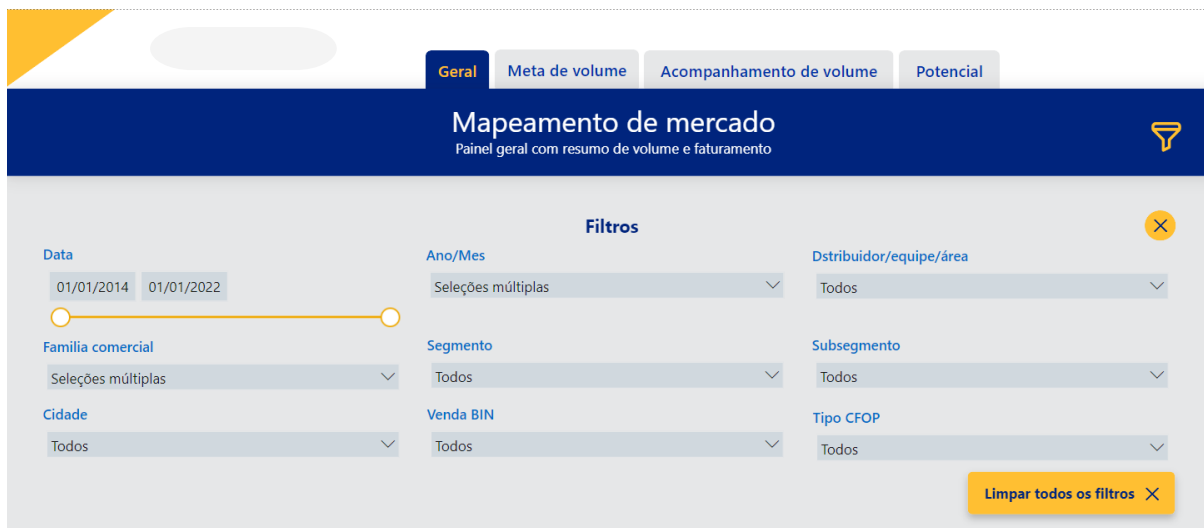
O autor deste documento foi responsável pelo desenvolvimento das visualizações descritas neste capítulo.

Figura 26 – Parte 3 da aba de acompanhamento detalhado de volume do painel Mapeamento de Mercado.



Fonte: autoria própria

Figura 27 – Seção de filtros de segmentação de dados.



Fonte: autoria própria

6 RESULTADOS FINAIS

Ao final do projeto, foi entregue um painel de *business intelligence* baseado em dados do setor comercial da empresa do ramo de manufatura de baterias elétricas, os quais estão formatados em um *data warehouse* central. Com isso, usuários de negócio da empresa em todas as unidades do país poderão se beneficiar do relatório para acompanhar se o volume de vendas de baterias está alcançando a meta estipulada.

A avaliação dos resultados obtidos perante os requisitos e critérios de sucesso definidos no capítulo 4 é abordada na seção a seguir.

6.1 ATENDIMENTO AOS REQUISITOS E CRITÉRIOS DE SUCESSO

6.1.1 Requisitos funcionais

Cumpriu-se todos os requisitos funcionais propostos, os quais estão atrelados às funcionalidades da entrega final.

6.1.2 Requisitos não funcionais

- Quanto ao desenvolvimento de relatórios de *business intelligence*:
 1. Performance: a escolha do método de conexão de dados *import*, no Power BI, foi preponderante na entrega de um relatório performático no quesito carregamento e interação de visuais. A modelagem de dados sob uma arquitetura *star-schema* também afetou positivamente a performance pela fato de haver poucas ligações de chaves entre as tabelas;
 2. Atualização diária: a configuração da atualização diária do relatório deve ser feita pela equipe da empresa contratante através do portal da inteligência, repositório central de painéis onde o mesmo será inserido;
 3. Dados históricos: adicionou-se, em todas as seções do relatório de mapeamento de mercado, gráficos que mostram a evolução histórica dos indicadores no tempo;
 4. Governança de dados: a técnica de filtragem de linhas de acordo com o usuário o qual acessa o relatório no Power BI foi executada através de conceitos de *row-level security* no Power BI. Entretanto, o uso de tal funcionalidade depende de ajustes no portal de inteligência os quais serão feitos após a escrita desta monografia por colaboradores da empresa contratante;
 5. Interface amigável: adotou-se a identidade visual da empresa contratante no relatório, com a escolha de cores padrões da empresa nos visuais;
 6. Navegação facilitada: inseriu-se botões de navegação na parte superior com o intuito de facilitar a troca de seção do relatório. Optou-se, também, pela

elaboração de seções mais cumpridas, permitindo que o usuário use a barra de rolagem para navegar pelos visuais. Por fim, os filtros foram inseridos em uma seção oculta acessível por um botão, atribuindo o foco do relatório para a visualização dos indicadores apenas.

- Quanto às ferramentas utilizadas:
 1. Power BI: utilizou-se o Power BI para o desenvolvimento do relatório de *business intelligence*;
 2. Microsoft Azure: fez-se o uso de funcionalidades da plataforma de computação na nuvem da Microsoft, como o Azure Repos e Azure Pipelines, ambos acessados através do Azure DevOps.

- Quanto à arquitetura de dados:
 1. Humanização do nome de tabelas e colunas: as colunas e tabelas receberam nomes que descrevem de forma clara o significado das mesmas. Os mesmos foram traduzidos para a língua portuguesa com o objetivo de garantir seus entendimento pelos usuários;
 2. Atualização diária dos dados: através de automações de tarefas no Azure Pipelines, os modelos do dbt são atualizados de forma diária baseados em novos dados extraídos no processo de EL (*Extract, Load*);
 3. Modelo de dados simplificado: os dados transformados estão dispostos em uma arquitetura baseada na metodologia *star-schema*, com a separação dos modelos em tabelas de fatos e dimensões. Tal divisão facilita o entendimento dos dados no contexto dos processos da empresa.

6.1.3 Critérios de sucesso

Todos os critérios de sucesso foram atingidos.

6.2 VALIDAÇÃO DOS DADOS

A validação dos dados transformados teve papel crucial na garantia de entrega de um relatório de *business intelligence* confiável e que suportasse as tomadas de decisões estratégicas da empresa do ramo de manufatura de baterias elétricas.

Fez-se, inicialmente, a validação dos dados através de comparações entre os valores dos indicadores no relatório já existente de mapeamento de mercado no portal de inteligência com o desenvolvido pelo autor. Foram percebidas divergências entre ambos, fato que exigiu retrabalho na transformação de dados. Notou-se, também, que as extrações incrementais diárias configuradas pelo time de engenheiros de dados

da Indicium estavam causando certa defasagem nos dados. Logo, após mudança de horários das mesmas, o processo foi normalizado.

Finalmente, após os eventuais ajustes na transformação de dados e na extração, os dados foram validados integralmente, garantindo que os usuários da empresa contratante tenham acesso a indicadores chaves confiáveis.

7 CONCLUSÃO

Conclui-se, com o final da implementação do projeto descrito nesta monografia, que os resultados obtidos satisfazem os requisitos propostos. Foi entregue uma estrutura de dados em um *data warehouse*, o qual desempenhará papel de fonte única de verdade na empresa. As tabelas estruturadas no DW alimentam painéis de *business intelligence*, os quais também estavam englobados no escopo do projeto.

Os dados, os quais foram transformados seguindo boas práticas de modelagem dimensional, estão estruturados em uma arquitetura a qual permite a fácil implementação dos mesmos em ferramentas de *business intelligence*. A escolha da aplicação dbt para a realização da transformação dos dados, aliada com sistemas de versionamento de código e computação na nuvem, torna o projeto escalável na medida em que cria uma estrutura replicável para eventuais novas fontes de dados.

Da mesma forma, a criação de um relatório de *business intelligence* performático, o qual suporta tomadas de decisões dos usuários de negócio seguindo boas práticas de *storytelling* e atendendo à identidade visual da empresa do ramo de manufatura de baterias elétricas, é um passo considerável rumo à padronização no desenvolvimento de painéis na empresa.

7.1 PROPOSTA DE TRABALHOS FUTUROS

A partir de dados tratados e validados, torna-se possível a criação de novas visões as quais auxiliam os usuários a ter um panorama mais completo da situação da empresa.

A implementação do indicador de taxa de *churn* de clientes possibilitaria mapear a porcentagem de clientes que deixaram de comprar produtos da empresa contratante. O KPI poderia ser agregado pelos diferentes atributos das dimensões, como cidade e unidades de distribuidores. Assim, seria possível identificar qual região tem sofrido maior abandono de clientes, viabilizando que maiores investimentos em campanhas de *marketing* fossem destinados à elas, por exemplo.

Por fim, a classificação de clientes em grupos distintos de acordo com a frequência de compra de baterias, *ticket* médio e data de última compra é atribuída através da técnica RFV (iniciais de Recência, Frequência, Valor). Cada cliente é inserido em um grupo de acordo com a pontuação que o mesmo atinge em termos dos KPIs citados. Logo, por abranger três indicadores relevantes, a técnica RFV pode fornecer *insights* precisos para diferenciar clientes de acordo com a importância dos mesmos para a empresa.

REFERÊNCIAS

- ATLASSIAN.COM. **Fluxo de trabalho de Gitflow**. 2022. Disponível em: <https://www.atlassian.com/br/git/tutorials/comparing-workflows/gitflow-workflow>.
- BHATTI, M.; AWAN, Hassan; RAZAQ, Z. The key performance indicators (KPIs) and their impact on overall organizational performance. **Quality Quantity**, v. 48, nov. 2014. DOI: 10.1007/s11135-013-9945-y.
- BLASI, Isabela. **A Jornada Data Driven para Maturidade Analítica**. 2021. Disponível em: <https://blog.indicium.tech/jornada-data-driven/>.
- CARROLL, Claire. **How we structure our dbt projects**. 2019. Disponível em: <https://discourse.getdbt.com/t/how-we-structure-our-dbt-projects/355>.
- CHAN, Frank K.Y.; THONG, James Y.L. Acceptance of agile methodologies: A critical review and conceptual framework. **Decision Support Systems**, v. 46, n. 4, p. 803–814, 2009. IT Decisions in Organizations. ISSN 0167-9236. DOI: <https://doi.org/10.1016/j.dss.2008.11.009>. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0167923608002133>.
- CHAUDHURI, Surajit; DAYAL, Umeshwar; NARASAYYA, Vivek. An Overview of Business Intelligence Technology. **Commun. ACM**, Association for Computing Machinery, New York, NY, USA, v. 54, n. 8, p. 88–98, ago. 2011. ISSN 0001-0782. DOI: 10.1145/1978542.1978562. Disponível em: <https://doi.org/10.1145/1978542.1978562>.
- CHOMAL, Vikas; SAINI, Jatinderkumar. Software Project Documentation – An Essence of Software Development. **International Journal of Advanced Networking and Applications**, v. 6, p. 2563–2572, mai. 2015.
- DATA type. 2016. Disponível em: <https://www.techtarget.com/searchapparchitecture/definition/data-type>.
- DOLENCE, Michael G.; NORRIS, Donald M. Using key performance indicators to drive strategic decision making. **New Directions for Institutional Research**, v. 1994, p. 63–80, 1994.

FAROOQ, Mubashar; KHAN, Dr. Impact of Training and Feedback on Employee Performance. **Far East Journal of Psychology and Business**, 5 No 1 Paper 2 October, p. 23–33, jan. 2011.

GARTNER.COM. **Analytics and Business Intelligence (ABI)**. 2022. Disponível em: <https://www.gartner.com/en/information-technology/glossary/business-intelligence-bi>.

JENSEN, Kenneth. **Cross-industry standard process for data mining**. 2012. Disponível em: https://en.wikipedia.org/wiki/Cross-industry_standard_process_for_data_mining.

KIMBALL, Ralph; ROSS, Margy. **The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling**. 3. ed. Indianapolis, IN: Wiley, 2013.

LABS, dbt. **The analytics engineering workflow**. 2022. Disponível em: <https://www.getdbt.com/>.

MEYER, Alex. **ERD: Retail, RTT**. 2003. Disponível em: <https://alexdmeyer.com/ax2012erd/Erd-One-Retail-R1-RTT.htm>.

MICROSOFT. **Entenda o esquema em estrela e a importância para o Power BI**. 2022. Disponível em: <https://docs.microsoft.com/pt-br/power-bi/guidance/star-schema>.

MICROSOFT.COM. **O que é Power BI?** 2022. Disponível em: <https://docs.microsoft.com/pt-br/power-bi/fundamentals/power-bi-overview>.

NERUR, Sridhar; MAHAPATRA, RadhaKanta; MANGALARAJ, George. Challenges of Migrating to Agile Methodologies. **Commun. ACM**, Association for Computing Machinery, New York, NY, USA, v. 48, n. 5, p. 72–78, mai. 2005. ISSN 0001-0782. DOI: 10.1145/1060710.1060712. Disponível em: <https://doi.org/10.1145/1060710.1060712>.

OMAH, Okechukwu; OBIEKWE, Onyebuchi. Impact of Employee Job Satisfaction on Organizational Performance. **International Journal of Current Research**, v. 6, p. 2343–403, dez. 2019.

PERVEEZ, Sayeda Haifa. **What is Git: Features, Command and Workflow in Git.** 2022. Disponível em:

<https://www.simplilearn.com/tutorials/git-tutorial/what-is-git>.

RAHM, Erhard; DO, Hong. Data Cleaning: Problems and Current Approaches. **IEEE Data Eng. Bull.**, v. 23, p. 3–13, jan. 2000.

SCHRÖER, Christoph; KRUSE, Felix; GÓMEZ, Jorge Marx. A Systematic Literature Review on Applying CRISP-DM Process Model. **Procedia Computer Science**, v. 181, p. 526–534, 2021. CENTERIS 2020 - International Conference on ENTERprise Information Systems / ProjMAN 2020 - International Conference on Project MANagement / HCist 2020 - International Conference on Health and Social Care Information Systems and Technologies 2020, CENTERIS/ProjMAN/HCist 2020. ISSN 1877-0509. DOI: <https://doi.org/10.1016/j.procs.2021.01.199>. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1877050921002416>.

SCRUM.ORG. **A Better Way Of Building Products.** 2022. Disponível em:

<https://www.scrum.org/resources/what-is-scrum>.

ANEXO A – TECNOLOGIAS UTILIZADAS

Figura 28 – Conjunto de tecnologias englobadas em todas as frentes do projeto.

