



UNIVERSIDADE FEDERAL DE UBERLÂNDIA  
BACHARELADO EM GESTÃO DA INFORMAÇÃO

MURILLO ANDRADE FERREIRA

**IMPLANTAÇÃO DE UM *DATA WAREHOUSE* COM A FERRAMENTA GAIO**

UBERLÂNDIA

2022

MURILLO ANDRADE FERREIRA

**IMPLANTAÇÃO DE UM *DATA WAREHOUSE* COM A FERRAMENTA GAIO**

Monografia apresentada ao Curso de Graduação em Gestão da Informação, da Universidade Federal de Uberlândia, como exigência parcial para a obtenção do título de Bacharel.

Orientador: Prof. José Eduardo Ferreira Lopes.

UBERLÂNDIA

2022

# IMPLANTAÇÃO DE UM *DATA WAREHOUSE* COM A FERRAMENTA GAIO

Murillo Andrade Ferreira

Monografia aprovada para a obtenção do título de Bacharel no Curso de Graduação em Gestão da Informação da Universidade Federal de Uberlândia (MG) pela banca examinadora formada pelos seguintes professores:

---

**Prof. José Eduardo Ferreira Lopes**

Professor Orientador

---

**Prof. Peterson Gandolfi**

Banca Examinadora

---

**Prof. Antônio Penedo**

Banca Examinadora

Uberlândia, 11 de agosto de 2022.

## RESUMO

O objetivo deste relato tecnológico é descrever e apresentar o processo realizado na construção de um *Data Warehouse* para o armazenamento e entrega de dados que caracterizam, com grande detalhamento e integridade, as vendas de uma empresa do mercado de proteção veicular, possibilitando assim a entrega, durante o dia a dia da empresa, de informações capazes de ajudar seus gestores a traçar novas metas e planos como também a se programarem para possíveis crises previstas. Após a implantação do *Data Warehouse*, foi descrita a evolução da qualidade dos dados. As conclusões foram que, com a implantação, o *Data Warehouse* se torna importante para o futuro da organização, que possui o desejo de guiar as decisões de seus vários departamentos com uma forte orientação de seus dados.

**Palavras-Chave:** *Data Warehouse*, Decisões, Dados, ETL e Integridade

## **ABSTRACT**

The main goal of this technical report is to describe and present the process used in a Data Warehouse creation to deliver, which aims to deliver righteous and detailed data from the sales of a company in the vehicle protection market, making it possible to provide daily information to help managers setting goals and plans, as well as get prepared for predicted crisis. After the Data Warehouse's implantation, was described the evolution of the data quality. The conclusions were that, with the Data Warehouse implantation, it becomes indispensable for the organization's future, which aims to guide multiple departments decisions heavily based on their data.

**Keywords:** Data Warehouse, Decisions, Data, ETL and Integrity

## LISTA DE ABREVIATURAS E DE SIGLAS

BI	<i>Business Intelligence</i>
DW	<i>Data Warehouse</i>
ETL	<i>Extract Transform Load</i>
SQL	<i>Structured Query Language</i>
B2B	<i>Business-to-Business</i>
B2C	<i>Business-to-Consumer</i>

## LISTA DE FIGURAS

Figura 1 <i>Framekework</i> de um processo de ETL.....	11
Figura 2 Canvas Inicial do Gaio .....	16
Figura 3 Demonstração de Código SQL Utilizado Dentro de uma Consulta .....	17
Figura 4 Canvas do Processo de Cruzamento e Tratamento de Dados .....	18
Figura 5 Primeira Transformação.....	19
Figura 6 Segunda Transformação.....	20
Figura 7 Cruzamento dos Dados.....	20
Figura 8 Canvas do Processo de Criação de Parâmetros .....	21
Figura 9 Canvas do Último Processo do Projeto.....	22
Figura 10 Ilustração da Organização dos Dados no Banco Operacional .....	23
Figura 11 Ilustração da Organização dos Dados no Data Warehouse.....	23

## SUMÁRIO

1 INTRODUÇÃO.....	8
2 REFERENCIAL TEÓRICO.....	10
2.1. ETL, <i>Data Warehouse</i> e Linguagem SQL .....	10
3 SITUAÇÃO PROBLEMA.....	13
4 TIPO DE INTERVENÇÃO ADOTADA.....	14
5 RESULTADOS OBTIDOS E ANÁLISE.....	15
6 CONSIDERAÇÕES FINAIS.....	24
7 REFERÊNCIAS.....	25



## INTRODUÇÃO

Com o avanço e desenvolvimento acelerado da tecnologia, a sofisticação de todos os possíveis mercados também acompanharam essa tendência. Este cenário está exigindo fortes adaptações das empresas que almejam ou já obtêm sucesso em seus respectivos negócios.

Os consumidores já se acostumaram com o fácil acesso à informação, atualmente as decisões de compra destes são tomadas de forma rápida, podendo se decidirem em adquirir um produto ou serviço na empresa “X” ou “Y” em instantes.

As organizações que desejam se destacar em seus respectivos mercados possuem a necessidade e a obrigação de acompanhar esse rápido fluxo de informação. Para conseguirem tal feito, elas são muito dependentes de dados, sejam estes produzidos por elas mesmas ou de uma fonte externa. Com estes dados, o *Business Intelligence* (BI) pode ser implantado nestas empresas e a corrida em busca da vantagem competitiva começa.

“O BI ajuda as organizações a analisar dados históricos e atuais, para que possam descobrir rapidamente insights práticos para a tomada de decisões estratégicas.” (MICROSOFT, 2022)

Conforme Silva e Terra (2015),

Saber usar essa ferramenta para acumular e separar as informações detidas pela empresa sobre seus processos e resultados pode acarretar em grandes mudanças para a empresa, pois a mesma terá maior *Know how* de mercado e poderá tomar as decisões estratégicas a seu favor, melhorando nas áreas de atuação que não trazem atualmente bons resultados e investindo ainda mais nas áreas que trazem grandes resultados e são fatores importantes para ganhar a confiança do cliente e conquistar a atenção de possíveis clientes, trazendo assim resultados ainda melhores visando sempre a maximização dos lucros(SILVA; TERRA, 2015).

Segundo Turban et al. (2009, p.27) o processo de *Business Intelligence* “constitui de bases de dados, metodologias, arquiteturas, canais e softwares que promovem o processo de transformar os dados em informações, decisões e, por fim, práticas necessárias ao negócio”.

Para uma solução de *Business Intelligence* ser devidamente implementada em uma organização e ela consiga aproveitar todos os seus benefícios é necessária uma estrutura tecnológica com bancos de dados e *softwares* que auxiliem o profissional da área a desempenhar funções importantes para a efetivação e

finalização de um projeto de BI.

Para que ocorra a viabilidade de entrega de informações corretas e valiosas para os gestores de empresas tomarem decisões que definirão o futuro do negócio é necessário garantir a integridade dos dados, evento que é validado com um fluxo de tarefas que se sucedem até o fim do projeto e possuem como finalidade transformar dados brutos ou inconsistentes em informações.

Durante a rotina de uma empresa, a todo momento ocorre a inserção, atualização e remoção de dados, sejam eles antigos ou mais recentes, em um banco de dados operacional, tudo depende das características operacionais e do modelo de negócio da organização.

Para que não haja transtornos como lentidão ou travamento do sistema, fato que proporciona uma experiência ruim para o cliente, e finalmente a aplicação do conceito de *Business Intelligence* comece, é necessário que se faça a extração de todos os dados que se encontram no banco operacional para uma *stage area*, que é um banco de dados provisório onde será desenvolvido toda a manipulação, tratamento e limpeza dos dados.

A partir do momento em que os dados estão adequados para deixarem a *stage area* eles são carregados em um *Data Warehouse (DW)*, acontecimento que marca o fim do processo de Extração, Transformação e Carga (ETL) no fluxo de tarefas. O DW é um grande banco de dados arquitetado, por um profissional de BI, de acordo com as necessidades e característica de cada negócio a fim de armazenar dados atuais e históricos da empresa otimizando assim o processo de consultas e análises de grandes volumes de dados. É por causa dele que é possível a análise de acontecimentos passados, como vendas, recebimentos, pagamentos e demissões, com um máximo de características descritivas possíveis, fato que possibilita tomadas de decisões precisas.

Um projeto chega ao seu final quando é conectado ao DW ferramentas de visualização de dados como o *Power BI* e *Tableau*, as mais famosas do mercado atual, onde será possível criar gráficos e *dashboards* que entreguem para os gestores das organizações os *insights* mais importantes. Esta é a única camada de todo o fluxo de tarefas que possui contato com o usuário final.

Assim, objetiva-se com este relato tecnológico, descrever com detalhamento as etapas realizadas nos processos de implantação de ETL e *Data Warehousing* desenvolvidas em um projeto real em um grupo empresarial.

## 1 REFERENCIALTEÓRICO

Para maior compreensão sobre o assunto é necessário conhecer melhor alguns conceitos importantes presentes em um projeto de *Business Intelligence*, focados em ETL (*Extract, Transform, Load*) e *Datawarehouse*.

### 1.1 ETL e Datawarehouse

Segundo Kimball e Caserta (2004) o processo de ETL é a base do *Data Warehouse*, pois este processo, quando devidamente projetado, é responsável pela extração dos dados nos sistemas de origem, imposição de padrões de qualidade e consistência e a oferta dos dados no formato próprio para os softwares com o objetivo de auxílio no processo de tomada de decisão.

Segundo Barbieri (2001), o conceito do processo de ETL pode ser dividido em cinco partes; filtro, integração, condensação, conversão e derivação de dados. O filtro de dados explica a tarefa de extração, momento em que se inicia o ETL e possui como função filtrar os dados que possam ter alguma característica indesejada que foram geradas nos bancos de dados operacionais comuns. A integração de dados é o passo onde se inicia a transformação e é o momento onde todos os dados começam a ser correlacionados por assunto, independente se as fontes que os derivaram são distintas. Na condensação de dados é onde os respectivos são organizados, por exemplo, os dados sobre as vendas passadas de uma empresa precisam estar juntos dos dados que se referem às vendas atuais. A conversão de dados tem como função a padronização destes em relação ao banco de dados utilizado no projeto, visto que cada banco apresenta dados de forma diferente, um exemplo desta situação é a apresentação de valores decimais, que em alguns ambientes podem ser separados por ponto e em outros por vírgula. Para finalizar o momento da transformação, a derivação de dados é a continuação da conversão, mas atuando apenas nas informações e não em como o modelo é empregado. Depois de desempenhar todas essas partes, os dados estão prontos para serem carregados no DW.

O *Framework* abaixo ilustra o processo de ETL.

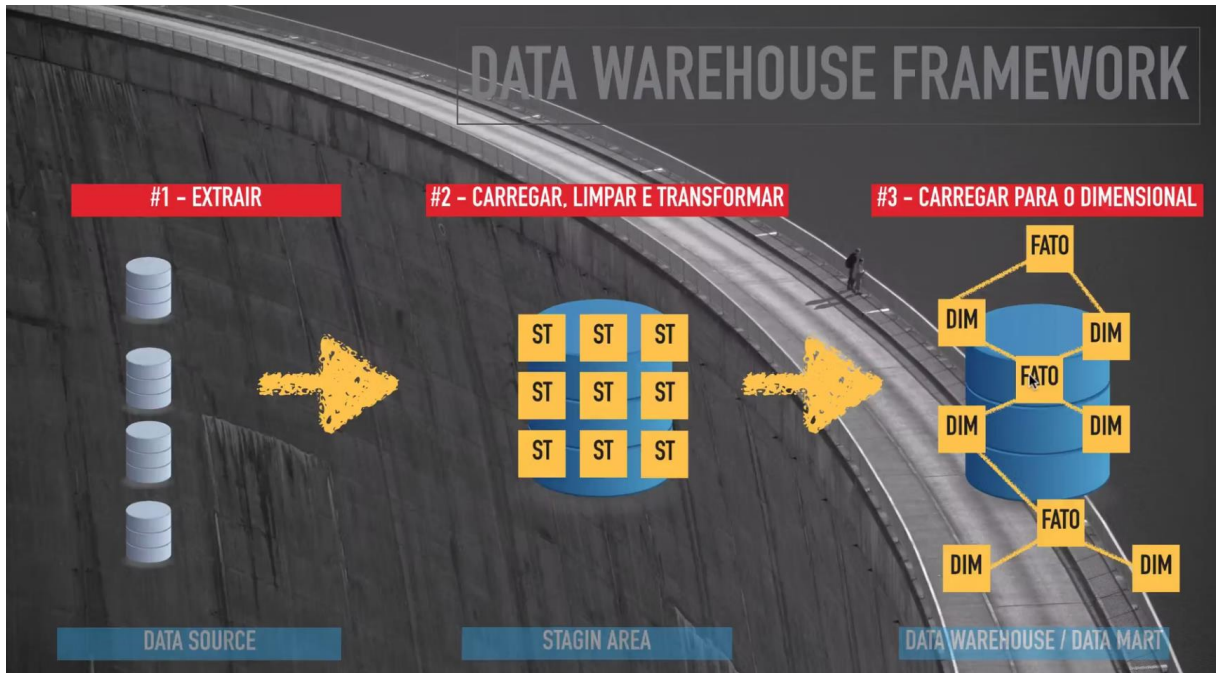


Figura 1: *Framekework* de um processo de ETL.  
Fonte: Adaptado de Piton (2016).

Uma definição encontrada em GUPTA(1997), define data warehouse como um ambiente estruturado e extensível projetado para o trabalho de análise de dados não voláteis, lógica e fisicamente transformados oriundos de diferentes fontes e alinhados como os objetivos estratégicos da empresa.

Para Leme (2004), DW é uma das partes fundamentais para que um projeto de BI seja executado.

O *Data Warehouse* é uma estrutura de banco de dados, ou ainda um repositório de dados que utiliza um sistema gerenciador de banco de dados com a finalidade de armazenar os dados tratados oriundos de diferentes fontes de dados (TURBAN et al., 2009).

O DW é quem centraliza os dados da empresa e elimina os ruídos de comunicação entre os departamentos, deixando tudo unificado. O *Data Warehouse* é todo planejado com foco na tomada de decisão e no que o negócio precisa (PITON, 2016).

Segundo Piton (2016), um DW desempenha as seguintes tarefas:

- “Apresenta e analisa eventos passados, podendo tomar decisões mais precisas com eles.
- Entrega as informações corretas para as pessoas certas no tempo ideal.

- Controla globalmente a visibilidade das informações e sabe onde estão sendo utilizados e por quem.
- Garante a unicidade de conceitos e regras de negócio e a confiabilidade dos dados entre todos os setores da empresa.
- Obtém respostas muito rápidas a consultas feitas em uma base com grande volume de dados.
- Como os dados agora estão centralizados, você gera relatórios com informações reais e atualizadas”.

À medida que os dados são derivados do ambiente operacional, passam por um processo de integração em que é preciso usar regras de conversão, ordenação e agregação, unificar entidades semelhantes, padronizar tamanhos, termos e valores etc. (BOHN,1997).

É por meio da integração dos dados e da orientação por assunto que o DW armazena dados de diferentes fontes de maneira consistente e condensada. Um tipo de integração a ser feita, por exemplo, se dá quando há diversidades entre nomenclaturas em colunas que se referenciam a um mesmo assunto ou ainda quando unidades de medidas que se encontram diferentes, mas precisam estar padronizadas (TURBAN et al., 2009).

## 2 SITUAÇÃO PROBLEMA

A organização pesquisada é um grupo empresarial que presta vários serviços como proteção veicular, assistência 24 horas e convênio de saúde e que fica situada em Uberlândia, Minas Gerais. A empresa, atualmente, está presente em 25 cidades do Brasil e atua tanto em B2B (cliente corporativo) quanto B2C (cliente varejo). Para o projeto de construção do DW foram analisados somente os dados referentes ao serviço de proteção veicular.

A organização fundada em 2013, conseguiu crescer relativamente rápido e pôde se expandir com sucesso. Com isso, cresceu também a necessidade da empresa em aumentar o seu autoconhecimento, nascendo assim um forte desejo em seus gestores de possuírem o conhecimento dos pontos fortes e fracos de seus respectivos departamentos.

A partir deste cenário a empresa tomou a decisão de implantar o *Business Intelligence* em seu negócio para deixar que seus próprios dados auxiliem os administradores da empresa a encontrarem respostas para seus questionamentos.

Logo, em um primeiro momento, foi notado pelos responsáveis pela implantação do projeto de BI que a organização sofria com um grande problema em guardar seus dados históricos devido as características de sua operação que requer frequentes atualizações em algumas características de seus dados, e do seu banco de dados operacional que não possui uma estrutura adequada que consiga preservar as características destes dados recém atualizados. Notou-se então, que a empresa perdia valiosas informações históricas.

Assim, o trabalho aqui apresentado tem como situação problema a necessidade de fornecer à organização uma solução que consiga proporcionar um ambiente pronto e preparado que consiga garantir a integridade de seus dados para que o projeto de BI possa entregar todas as análises desejadas pelos gestores de negócio.

### 3 TIPO DE INTERVENÇÃO ADOTADA

A intervenção para a situação problema é a construção de um *Data Warehouse* para que todos os dados históricos sobre as vendas de proteção veicular e seus respectivos clientes, estejam salvos e armazenados em um ambiente estruturado e propício à entrega de informações íntegras que apoiarão fortemente os tomadores de decisão da empresa.

O Grupo optou pela contratação da Gaio Innotech como fornecedora da ferramenta que foi utilizada para a realização do projeto de *Business Intelligence*. A solução Gaio, desenvolvido também na cidade de Uberlândia, consegue entregar todas as etapas de um projeto de BI, de ponta a ponta, desde as extrações até as visualizações de dados.

A construção do DW foi designada ao estagiário de BI do Grupo, que foi responsável por todo o *back-end* dos projetos de análise de dados da empresa. O estagiário tinha como suporte o seu gestor de departamento, no qual oferecia suporte técnico e esclarecia dúvidas sobre como funcionava as regras de negócios e as atividades operacionais da empresa. Diante desta situação o estagiário tinha condições favoráveis para a realização das tarefas. Além do suporte de negócios, o estagiário contou também com o suporte técnico da equipe Gaio.

O projeto foi desenvolvido em três semanas, na qual a primeira teve seu maior tempo reservado para reuniões e conversas entre os envolvidos para o alinhamento dos problemas e ajustes de expectativas. As semanas restantes foram designadas para o desenvolvimento técnico. Assim, a construção do DW foi concluída e está sendo frequentemente monitorada pelos responsáveis para que haja a identificação de possíveis inconsistências e também, para que aconteça a verificação da entrega de informações corretas, garantindo assim um sistema com característica íntegra.

## 4 RESULTADOS OBTIDOS E ANÁLISE

Uma parte importante na engrenagem de um projeto de BI é a linguagem de manipulação de dados denominada Structured Query Language (SQL). O SQL é a única e exclusiva linguagem dos bancos de dados relacionais, só é possível realizar as diversas tarefas necessárias dentro deles depois de entender a lógica e os padrões da linguagem. Segundo Costa (2007), os comandos SQL podem ser divididos em sub-linguagens, como a DML (*Data Manipulation Language*) que é a responsável pelas manipulações dos dados por meio de seleções, inclusões, alterações e exclusões dentro das tabelas e a DDL (*Data Definition Language*) que são as listas de comandos que criam e fornecem manutenção às estruturas e objetos do banco de dados, como por exemplo as tabelas.

Partindo da aplicação já finalizada, a primeira tela do Gaio, vista na visão do desenvolvedor *back-end*, é dividida entre um grande canvas, onde é mostrado visualmente todos os processos realizados, e um espaço à esquerda onde é possível a navegação entre processos desenvolvidos e das tarefas disponíveis, como a criação de parâmetros para os dados, administração dos bancos de dados conectados e o *input* de planilhas externas.

Na Figura 2 é possível visualizar o canvas do processo de extração de dados, à direita. Nele, estão desenhadas todas as ações de extração de dados do banco de dados operacional do Grupo. As miniaturas em azul com seus respectivos nomes em baixo, representam as tabelas do banco de dados operacional, escolhidas como fonte de dados. As miniaturas pretas representam as consultas feitas na linguagem SQL. E por fim, as miniaturas vermelhas representam tabelas contendo os dados resultantes de suas respectivas consultas.



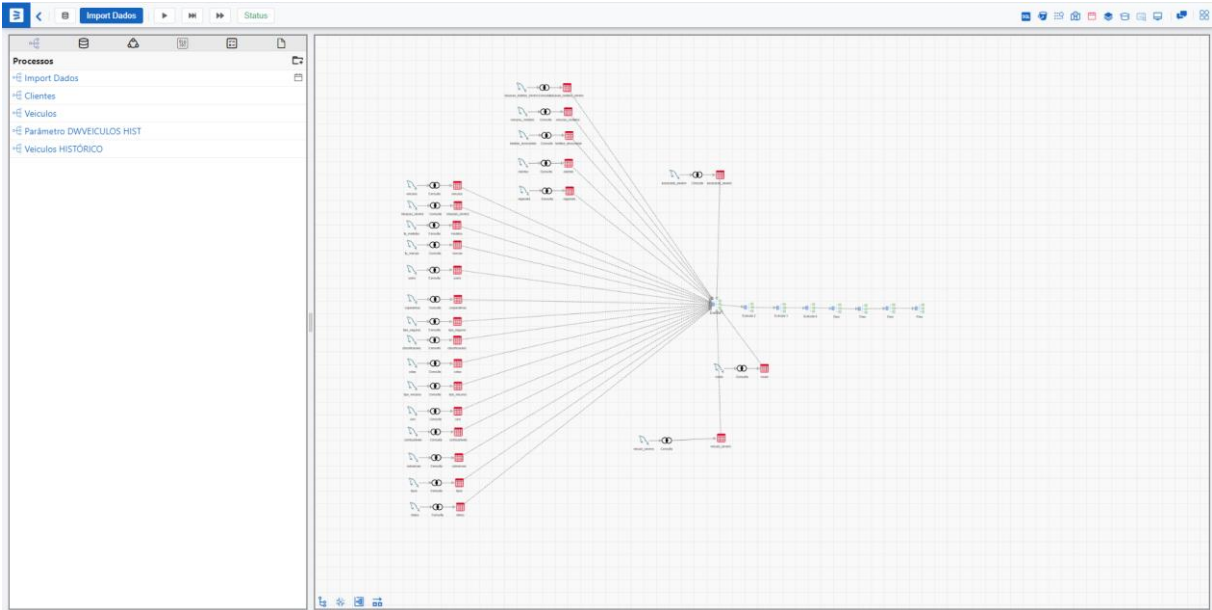


Figura 2: Canvas Inicial do Gaio.  
Fonte: Empresa estudada.

Neste primeiro processo de extração de dados é onde começa o ETL visando a criação do DW. Todas as consultas são para extrair os dados dos bancos de dados fonte e colocá-los em tabelas que ficarão armazenadas dentro do repositório do Gaio, que é o banco de dados dedicado para a aplicação.

A Figura 3 ilustra um exemplo de parte do conteúdo presente dentro de uma das consultas. Neste caso, com a linguagem SQL, são selecionados e transferidos todos os dados da tabela do banco de dados operacional nomeada de “veiculos” para uma tabela provisória dentro do banco de dados dedicado à aplicação.

```

1 SELECT
2 `veiculos`.`id`,
3 `veiculos`.`id_voluntario`,
4 `veiculos`.`id_coperativa`,
5 `veiculos`.`tp_seguro`,
6 `veiculos`.`recorrencia`,
7 `veiculos`.`id_classificacao`,
8 `veiculos`.`id_cota`,
9 `veiculos`.`tipo_veiculo`,
10 `veiculos`.`id_cor`,
11 `veiculos`.`id_combustivel`,
12 `veiculos`.`tp_cobranca`,
13 `veiculos`.`id_tipo`,
14 `veiculos`.`dt_vencimento`,
15 `veiculos`.`dt_contrato`,
16 `veiculos`.`vl_adesao`,
17 `veiculos`.`id_externo`,
18 `veiculos`.`vl_mensal`,
19 `veiculos`.`txtobs`,
20 `veiculos`.`marca`,
21 `veiculos`.`modelo`,
22 `veiculos`.`ano_fb`,
23 `veiculos`.`ano_modelo`,
24 `veiculos`.`cod_fipe`,
25 `veiculos`.`valor`,
26 `veiculos`.`dt_reativacao`,
27 `veiculos`.`chassi`,
28 `veiculos`.`horario`,
29 `veiculos`.`renavam`,
30 `veiculos`.`placa`,
31 `veiculos`.`dt_vistoria`,
32 `veiculos`.`id_cliente`,
33 `veiculos`.`p_premio`,
34 `veiculos`.`nosso_numero`,
35 `veiculos`.`km5001`,
36 `veiculos`.`boleto_interno`,
37 `veiculos`.`km5002`,
38 `veiculos`.`km700`,
39 `veiculos`.`km1000`,
40 `veiculos`.`chkdanos`,
41 `veiculos`.`copr1`,
42 `veiculos`.`providros`,
43 `veiculos`.`copr2`,
44 `veiculos`.`uber`,
45 `veiculos`.`chkdanos1`,
46 `veiculos`.`rastreadorchk`,
47 `veiculos`.`carres`,
48 `veiculos`.`id_status`,
49 `veiculos`.`token`,
50 `veiculos`.`id_situacao`,
51 `veiculos`.`id_vistoriador`,
52 `veiculos`.`id_parceiro`,
53 `veiculos`.`created_at`,
54 `veiculos`.`updated_at`,
55 `veiculos`.`id_captacao`,
56 `veiculos`.`log_erro_cadastro`,
57 `veiculos`.`mototaxi`,
58 `veiculos`.`financiado`
59 FROM
60 `seven_db`.`veiculos`

```

Figura 3: Demonstração de Código SQL Utilizado Dentro de uma Consulta.  
Fonte: Empresa estudada.

Esse foi o primeiro passo técnico do projeto, onde foi realizado um processo semelhante ao da Figura 3 para cada tabela que possuía dados julgados necessários e importantes para o negócio do Grupo. Ao final do processo tem-se como resultado a extração completa destes dados para um ambiente adequado, possibilitando assim o começo das transformações, análises e cruzamentos.

A Figura 4 representa o canvas do processo onde foram cruzado todos os dados necessários para a construção do DW de vendas. É possível observar que nele estão presentes um número menor de consultas em relação ao processo passado, isso se deve ao fato de que neste momento não está sendo extraído dados de suas respectivas fontes do banco operacional para a sua fonte correspondente no

banco de dados do Gaio. O processo representado pela Figura 4 está descrevendo duas transformações e um cruzamento geral de dados.

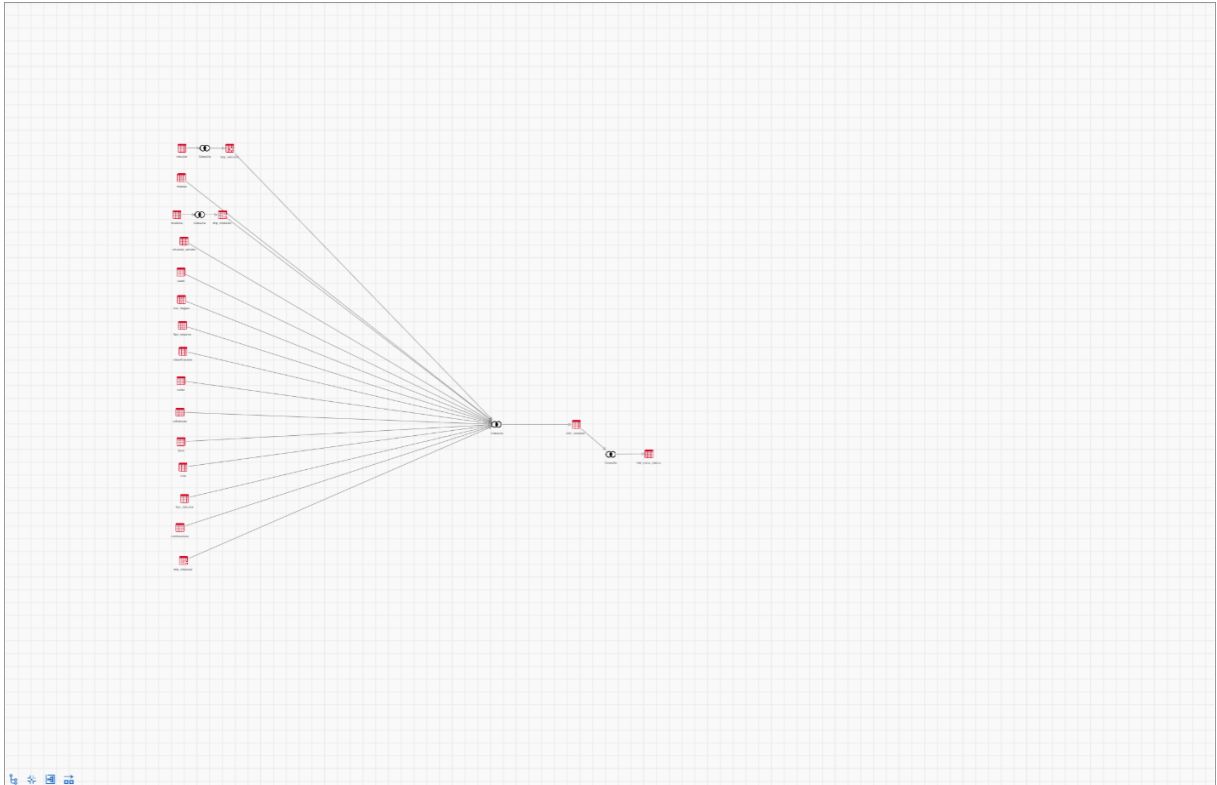


Figura 4: Canvas do Processo de Cruzamento e Tratamento de Dados.  
Fonte: Empresa estudada.

A Figura 5 demonstra parte dos códigos SQL utilizados na primeira transformação de dados dentro da consulta. Na fonte de dados primária, no banco de dados operacional, a venda possui apenas o dado de data da venda para caracterizar o tempo, os gestores do negócio relataram que seria muito importante para eles a visualização do tempo da venda descrita por ano comercial e ciclo comercial, duas nomenclaturas que caracterizam o período de venda da empresa. Assim, os códigos SQL, utilizando “*case-when*” foram criados para transformar o dado da data da venda, sem excluí-lo, em ano comercial e ciclo comercial possibilitando uma análise mais profunda em relação ao tempo da venda.

```

1 case
2
3 when (dt_contrato) between '2012-12-21' and '2013-01-20' then '2013'
4 when (dt_contrato) between '2013-01-21' and '2013-02-20' then '2013'
5 when (dt_contrato) between '2013-02-21' and '2013-03-20' then '2013'
6 when (dt_contrato) between '2013-03-21' and '2013-04-20' then '2013'
7 when (dt_contrato) between '2013-04-21' and '2013-05-20' then '2013'
8 when (dt_contrato) between '2013-05-21' and '2013-06-20' then '2013'
9 when (dt_contrato) between '2013-06-21' and '2013-07-20' then '2013'
10 when (dt_contrato) between '2013-07-21' and '2013-08-20' then '2013'
11 when (dt_contrato) between '2013-08-21' and '2013-09-20' then '2013'
12 when (dt_contrato) between '2013-09-21' and '2013-10-20' then '2013'
13 when (dt_contrato) between '2013-10-21' and '2013-11-20' then '2013'
14 when (dt_contrato) between '2013-11-21' and '2013-12-20' then '2013'
15
16 when (dt_contrato) between '2013-12-21' and '2014-01-20' then '2014'
17 when (dt_contrato) between '2014-01-21' and '2014-02-20' then '2014'
18 when (dt_contrato) between '2014-02-21' and '2014-03-20' then '2014'
19 when (dt_contrato) between '2014-03-21' and '2014-04-20' then '2014'
20 when (dt_contrato) between '2014-04-21' and '2014-05-20' then '2014'
21 when (dt_contrato) between '2014-05-21' and '2014-06-20' then '2014'
22 when (dt_contrato) between '2014-06-21' and '2014-07-20' then '2014'
23 when (dt_contrato) between '2014-07-21' and '2014-08-20' then '2014'
24 when (dt_contrato) between '2014-08-21' and '2014-09-20' then '2014'
25 when (dt_contrato) between '2014-09-21' and '2014-10-20' then '2014'
26 when (dt_contrato) between '2014-10-21' and '2014-11-20' then '2014'
27 when (dt_contrato) between '2014-11-21' and '2014-12-20' then '2014'
28
29 when (dt_contrato) between '2014-12-21' and '2015-01-20' then '2015'
30 when (dt_contrato) between '2015-01-21' and '2015-02-20' then '2015'
31 when (dt_contrato) between '2015-02-21' and '2015-03-20' then '2015'
32 when (dt_contrato) between '2015-03-21' and '2015-04-20' then '2015'
33 when (dt_contrato) between '2015-04-21' and '2015-05-20' then '2015'
34 when (dt_contrato) between '2015-05-21' and '2015-06-20' then '2015'
35 when (dt_contrato) between '2015-06-21' and '2015-07-20' then '2015'
36 when (dt_contrato) between '2015-07-21' and '2015-08-20' then '2015'
37 when (dt_contrato) between '2015-08-21' and '2015-09-20' then '2015'
38 when (dt_contrato) between '2015-09-21' and '2015-10-20' then '2015'
39 when (dt_contrato) between '2015-10-21' and '2015-11-20' then '2015'
40 when (dt_contrato) between '2015-11-21' and '2015-12-20' then '2015'
41
42 when (dt_contrato) between '2015-12-21' and '2016-01-20' then '2016'
43 when (dt_contrato) between '2016-01-21' and '2016-02-20' then '2016'
44 when (dt_contrato) between '2016-02-21' and '2016-03-20' then '2016'
45 when (dt_contrato) between '2016-03-21' and '2016-04-20' then '2016'
46 when (dt_contrato) between '2016-04-21' and '2016-05-20' then '2016'
47 when (dt_contrato) between '2016-05-21' and '2016-06-20' then '2016'
48 when (dt_contrato) between '2016-06-21' and '2016-07-20' then '2016'
49 when (dt_contrato) between '2016-07-21' and '2016-08-20' then '2016'
50 when (dt_contrato) between '2016-08-21' and '2016-09-20' then '2016'
51 when (dt_contrato) between '2016-09-21' and '2016-10-20' then '2016'
52 when (dt_contrato) between '2016-10-21' and '2016-11-20' then '2016'
53
1 case
2 when toInt64(formatDateTime(dt_contrato, '%d')) between 21 and 31 and toInt64(formatDateTime(dt_contrato, '%m')) = 12
3 then 1
4 when toInt64(formatDateTime(dt_contrato, '%d')) between 1 and 20 and toInt64(formatDateTime(dt_contrato, '%m')) = 1
5 then 1
6 when toInt64(formatDateTime(dt_contrato, '%d')) between 21 and 31 and toInt64(formatDateTime(dt_contrato, '%m')) = 1
7 then 2
8 when toInt64(formatDateTime(dt_contrato, '%d')) between 1 and 20 and toInt64(formatDateTime(dt_contrato, '%m')) = 2
9 then 2
10 when toInt64(formatDateTime(dt_contrato, '%d')) between 21 and 31 and toInt64(formatDateTime(dt_contrato, '%m')) = 2
11 then 3
12 when toInt64(formatDateTime(dt_contrato, '%d')) between 1 and 20 and toInt64(formatDateTime(dt_contrato, '%m')) = 3
13 then 3
14 when toInt64(formatDateTime(dt_contrato, '%d')) between 21 and 31 and toInt64(formatDateTime(dt_contrato, '%m')) = 3
15 then 4
16 when toInt64(formatDateTime(dt_contrato, '%d')) between 1 and 20 and toInt64(formatDateTime(dt_contrato, '%m')) = 4
17 then 4
18 when toInt64(formatDateTime(dt_contrato, '%d')) between 21 and 31 and toInt64(formatDateTime(dt_contrato, '%m')) = 4
19 then 5
20 when toInt64(formatDateTime(dt_contrato, '%d')) between 1 and 20 and toInt64(formatDateTime(dt_contrato, '%m')) = 5
21 then 5
22 when toInt64(formatDateTime(dt_contrato, '%d')) between 21 and 31 and toInt64(formatDateTime(dt_contrato, '%m')) = 5
23 then 6
24 when toInt64(formatDateTime(dt_contrato, '%d')) between 1 and 20 and toInt64(formatDateTime(dt_contrato, '%m')) = 6
25 then 6
26 when toInt64(formatDateTime(dt_contrato, '%d')) between 21 and 31 and toInt64(formatDateTime(dt_contrato, '%m')) = 6
27 then 7
28 when toInt64(formatDateTime(dt_contrato, '%d')) between 1 and 20 and toInt64(formatDateTime(dt_contrato, '%m')) = 7
29 then 7
30 when toInt64(formatDateTime(dt_contrato, '%d')) between 21 and 31 and toInt64(formatDateTime(dt_contrato, '%m')) = 7
31 then 8
32 when toInt64(formatDateTime(dt_contrato, '%d')) between 1 and 20 and toInt64(formatDateTime(dt_contrato, '%m')) = 8
33 then 8
34 when toInt64(formatDateTime(dt_contrato, '%d')) between 21 and 31 and toInt64(formatDateTime(dt_contrato, '%m')) = 8
35 then 9
36 when toInt64(formatDateTime(dt_contrato, '%d')) between 1 and 20 and toInt64(formatDateTime(dt_contrato, '%m')) = 9
37 then 9
38 when toInt64(formatDateTime(dt_contrato, '%d')) between 21 and 31 and toInt64(formatDateTime(dt_contrato, '%m')) = 9
39 then 10
40 when toInt64(formatDateTime(dt_contrato, '%d')) between 1 and 20 and toInt64(formatDateTime(dt_contrato, '%m')) = 10
41 then 10
42 when toInt64(formatDateTime(dt_contrato, '%d')) between 21 and 31 and toInt64(formatDateTime(dt_contrato, '%m')) = 10
43 then 11
44 when toInt64(formatDateTime(dt_contrato, '%d')) between 1 and 20 and toInt64(formatDateTime(dt_contrato, '%m')) = 11
45 then 11
46 when toInt64(formatDateTime(dt_contrato, '%d')) between 21 and 31 and toInt64(formatDateTime(dt_contrato, '%m')) = 11
47 then 12
48 when toInt64(formatDateTime(dt_contrato, '%d')) between 1 and 20 and toInt64(formatDateTime(dt_contrato, '%m')) = 12
49 then 12
50 else null
51 end

```

Figura 5: Primeira Transformação.

Fonte: Criado pelo pesquisador.

A Figura 6 caracteriza quatro códigos SQL diferentes utilizados na segunda transformação. Os dados do banco operacional que descreviam o modelo do carro do cliente estavam totalmente inconsistentes, estes quatro códigos são os responsáveis por tratar as inconsistências e transformar os dados originais em dados confiáveis e íntegros para uma possível análise de vendas por modelo de carro.

```

1 case
2   when modelo like '% %' then mid(modelo, 1, position(modelo, ' '))
3   when modelo like '%' then modelo
4   end

mid(modelo, position(modelo, ' '), +100)

1 case
2   when resto_motor like '%1.%' then substring(resto_motor, position(resto_motor, '1.')-1, 4)
3   when resto_motor like '%2.%' then substring(resto_motor, position(resto_motor, '2.')-1, 4)
4   when resto_motor like '%3.%' then substring(resto_motor, position(resto_motor, '3.')-1, 4)
5   when resto_motor like '%4.%' then substring(resto_motor, position(resto_motor, '4.')-1, 4)
6   when resto_motor like '%5.%' then substring(resto_motor, position(resto_motor, '5.')-1, 4)
7   when resto_motor like '%6.%' then substring(resto_motor, position(resto_motor, '6.')-1, 4)
8   when resto_motor like '%7.%' then substring(resto_motor, position(resto_motor, '7.')-1, 4)
9   when resto_motor like '%8.%' then substring(resto_motor, position(resto_motor, '8.')-1, 4)
10  when resto_motor like '%9.%' then substring(resto_motor, position(resto_motor, '9.')-1, 4)
11  else resto_motor
12  end

concat(nome_carro, ' ', resto_motor2)

```

Figura 6: Segunda Transformação.  
 Fonte: Criada pelo pesquisador.

A Figura 7 demonstra visualmente a consulta onde todos os dados anteriores são cruzados. Para o cruzamento destes dados foi necessário uma análise muito cuidadosa e técnica por parte do estagiário, pois qualquer erro aqui poderia prejudicar a entrega íntegra de informações, neste passo é onde todos os dados são programados para se relacionarem entre si.

Figura 7: Cruzamento dos Dados.  
 Fonte: Empresa Estudada.

Esse cruzamento geral de dados é possível devido a característica da linguagem SQL com o comando “*join*”, nele as várias tabelas se relacionam por meio

de uma das colunas que possuem uma característica de dado que consegue diferenciar cada linha presente na respectiva tabela.

Ao final deste processo é gerada uma tabela final que possui todo o “esqueleto” do DW. Nela todos os dados de todas as tabelas escolhidas já estão tratados, transformados e cruzados, neste momento já é possível caracterizar as vendas feitas pelo Grupo Empresarial com um alto nível de detalhamento. O projeto não chegou ao seu fim ainda, O DW, até o momento, não é capaz de identificar alterações nos dados e guardar nele as informações destes dados antes e depois da alteração, a partir do momento em que isso é possível o DW estará apto a salvar dados históricos e pronto para o dia-a-dia da empresa.

A Figura 8 representa o canvas do processo onde são utilizadas as tarefas de desenvolvimento do esquema para a criação de parâmetros para os dados. A função dos parâmetros é, para cada nova requisição ao banco operacional que o DW fizer, ler todos os dados deste e indicar para o sistema quais deles tiveram alguma alteração ou então são de fato novos dados produzidos.

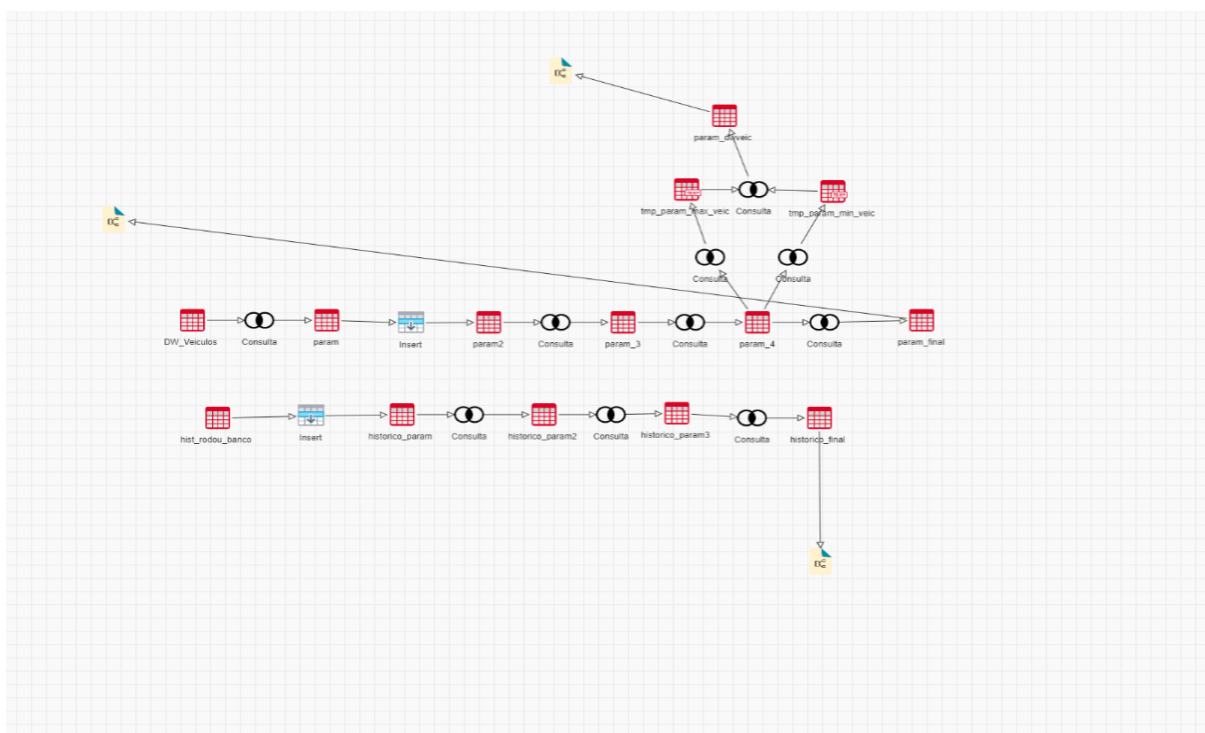


Figura 8: Canvas do Processo de Criação de Parâmetros.  
Fonte: Empresa Estudada.

A Figura 9 demonstra o último processo do projeto de construção do *Data Warehouse* utilizando a ferramenta Gaio. A partir do momento em que o DW é

avisado sobre quais dados foram alterados ou quais deles são novos, ele consegue filtrar todos estes com estas características e inseri-los aos demais dados antigos que já estavam armazenados. Agora, é possível caracterizar e diferenciar as vendas com um maior nível de detalhamento possível e com integridade.

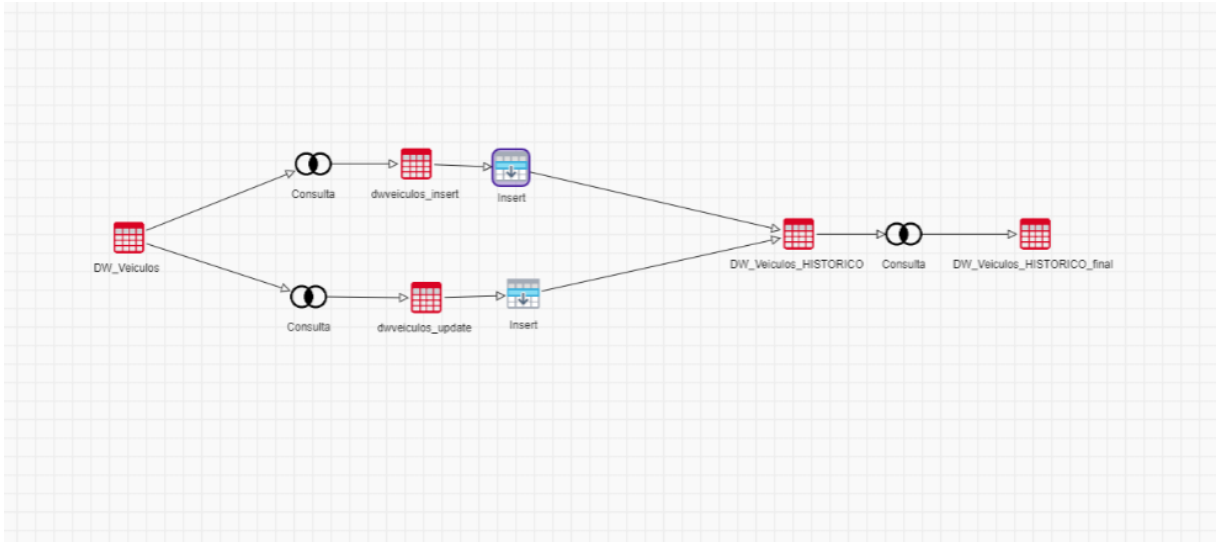


Figura 9: Canvas do Último Processo do Projeto.  
Fonte: Empresa Estudada.

Para uma maior ilustração do resultado final as Figuras 10 e 11 ilustram como o *Data Warehouse* entrega informações com alto valor de integridade e grande nível de detalhamento.

Foram selecionados apenas algumas colunas, e ainda com uma certa privacidade nos dados de placa, para servirem de exemplo do antes e depois do projeto. A Figura 10 ilustra como os dados são organizados no banco operacional, ou seja, no começo do projeto. Neste formato a entrega de informação é limitada e não oferece valor à tomada de decisão da empresa.

# id	dt_contrato	# id_cliente	placa	modelo	# id_voluntario	# id_coperativa	vl_adesao
1	2016-09-26	2751	[REDACTED]	2073	48	27	250,00
2	2019-03-21	2750	[REDACTED]	4135	44757	58	150
4	2016-09-26	1903	[REDACTED]	1104	162	61	250,00
5	2017-07-13	10308	[REDACTED]	606	168	61	250,00
6	2016-09-26	2747	[REDACTED]	2796	162	61	110,00
7	2016-09-23	2746	[REDACTED]	4421	168	61	250,00
9	2016-09-23	2744	[REDACTED]	2414	55	61	300
10	2016-09-23	2743	[REDACTED]	3968	48	27	300
11	2016-09-24	2742	[REDACTED]	3841	187	8	200,00
12	2016-09-24	2741	[REDACTED]	5465	119	61	250

Figura 10: Ilustração da Organização dos Dados no Banco Operacional.  
Fonte: Criada pelo Pesquisador.

A Figura 11 demonstra como os mesmos dados da Figura 10 são visualizados no *Data Warehouse*. A dificuldade do entendimento sobre o que aconteceu no fato da venda desaparece dando lugar ao entendimento completo da mesma, com os dados no mais alto nível de detalhamento e integração. Com este formato, os gestores de negócios encontram um ambiente favorável e preparado para guiá-los durante suas rotinas de tomadas de decisões.

# id	ciclo_comercial	ano_comercial	nome_cliente	placa	modelo	voluntario	cooperativa	vl_adesao
1	10	2016	[REDACTED]	[REDACTED]	Sc'nic 1.6	[REDACTED]	METTA	250,00
2	4	2019	[REDACTED]	[REDACTED]	Fiesta 1.0	[REDACTED]	CAÇADORES	150
4	10	2016	[REDACTED]	[REDACTED]	Monza 2.0	[REDACTED]	GLADIADORES	250,00
5	7	2017	[REDACTED]	[REDACTED]	Strada 1.8	[REDACTED]	GLADIADORES	250,00
6	10	2016	[REDACTED]	[REDACTED]	CG 150 TITAN-KS/ TITAN-JOB	[REDACTED]	GLADIADORES	110,00
7	10	2016	[REDACTED]	[REDACTED]	Punto 1.4	[REDACTED]	GLADIADORES	250,00
9	10	2016	[REDACTED]	[REDACTED]	Gol 1.0	[REDACTED]	GLADIADORES	300
10	10	2016	[REDACTED]	[REDACTED]	Astra 2.0	[REDACTED]	METTA	300
11	10	2016	[REDACTED]	[REDACTED]	BIZ 125 ES/ ES F.INJ./ES MIX F.INJECTION	[REDACTED]	CUABRASA	200,00
12	10	2016	[REDACTED]	[REDACTED]	UNO 1.4	[REDACTED]	GLADIADORES	250

Figura 11: Ilustração da Organização dos Dados no *Data Warehouse*.  
Fonte: Criada pelo Pesquisador.

Nesta estrutura, é possível destacar também como resultado o ganho em tempo na geração dos relatórios e informações, isso acontece devido a maneira de como os dados estão relacionados no DW, que é de forma desnormalizada, deixando assim o processo de tomada de decisão na empresa mais ágil.



## 5 CONSIDERAÇÕES FINAIS

Neste relato técnico, o objetivo foi apresentar o processo de construção de um *Data Warehouse*, dentro de um projeto de *Business Intelligence*, para exponenciar a qualidade, assertividade e velocidade das tomadas de decisões dentro de um grupo empresarial. A partir daqui é possível produzir gráficos e *dashboards* contendo visualmente todas as informações disponíveis dentro dos dados da empresa.

Com este poder, a organização pode se destacar dentro de seu mercado aumentando a sua competitividade, com todo este autoconhecimento é possível elaborar estratégias e planos para a empresa conseguir se destacar em relação aos seus concorrentes, podendo talvez até mesmo antecipar o futuro e se preparar para possíveis crises ou capturar oportunidades.

Cada empresa que decide adotar o BI possui um conjunto de ferramentas diferentes para a realização do projeto, seja por questões de preferência ou de viabilidade financeira. Devido a isso, o profissional de *Business Intelligence* deve sempre estar em contato e estudando os vários softwares do mercado. Todos eles realizam os processos de ETL e o planejamento e construção do DW de diferentes formas, assim o profissional precisa se atentar em dominar também o verdadeiro conceito destes processos e por isso essas são as tarefas que demandam maior conhecimento técnico

Este relato foi motivo de importância e orgulho para o até então estagiário. Com ele, foi possível um grande desenvolvimento técnico e teórico de suas habilidades na área de *Business Intelligence* e também um grande ganho de experiência em projetos de DW.

A partir daqui fica como uma de suas funções, além da participação em outros projetos, o monitoramento do funcionamento do *Data Warehouse* e fornecimento de suporte ao mesmo quando necessário, assim o Grupo se beneficiará sempre de informações de alto valor para tomadas de decisões.

## 6 REFERÊNCIAS

- BARBIERI, C. **BI - Business Intelligence – Modelagem e tecnologia**. Axcel Books, 2001.
- BOHN, K. **Converting Data for Warehouses**. DBMS Magazine. Junho, 1997.
- COSTA, R. L. DE C. **SQL Guia Prático**. Brasport Livros e Multimídia Ltda., 2007.
- GAIO. Gaio: Soluções de Big Data, 2021. Página inicial. Disponível em: <<https://gaio.io/>>. Acesso em: 20 de mar. de 2021.
- GUPTA, Vivek R., **An Introduction to Data Warehousing**. Disponível em: <[http://www.system-services.com.](http://www.system-services.com/)> Acesso em: 11 de jul. De 2022
- KIMBALL, Ralph; CASERTA, Joe. **The data warehouse ETL toolkit: practical techniques for extracting, cleaning, conforming and delivering data**. São Paulo: Wiley, 2004.
- O que é Business Intelligence**. Microsoft, 2022. Disponível em: <<https://powerbi.microsoft.com/pt-br/what-is-business-intelligence>>. Acesso em: 10 de jul. de 2022.
- LEME FILHO, Trajano. **Business intelligence no Microsoft Excel**. Rio de Janeiro: Axcel Books, 2004.
- PITON. (Brasil). *Data Warehouse*. In: PITON. (Brasil). **Data Warehouse – O que É?** [Porto Alegre, RS]: Piton. 2016. Disponível em: <https://rafaelpiton.com.br/blog/data-warehouse-o-que-e/>. Acesso em: 25 mar. 2021.
- PITON. (Brasil). *Data Warehouse*. In: PITON. (Brasil). **Data Warehouse – Para Que Serve?** [Porto Alegre, RS]: Piton. 2016. Disponível em <https://rafaelpiton.com.br/blog/data-warehouse-para-que-serve/>. Acesso em: 25 mar. 2021.
- SILVA, V. C. L.; TERRA, L. A. *Business Intelligence* como Fator Decisivo na Competitividade Empresarial: uma análise a partir de multicasos. **Revista Inteligência Competitiva.**, 2015. Disponível em <<http://www.inteligenciacompetitivarev.com.br/ojs/index.php/rev/article/view/107/pdf>>. Acesso em: 20 mar. 2021.
- TURBAN, E. et al. **Business intelligence: um enfoque gerencial para a inteligência do negócio**. [S.l.]: Bookman Editora, 2009.