



Universidade Federal do Rio de Janeiro

Escola Politécnica

MBA em Big Data, Business Intelligence e Business Analytics
(MB3B)

**O USO DE FERRAMENTAS DE BI NO TRATAMENTO E
VISUALIZAÇÃO DE DADOS EM UMA EMPRESA DE GRANDE PORTE**

Autor:

Caíque de Almeida Malafaia Souza

Orientador:

Manoel Villas Boas Junior, M. Sc.

Coorientador:

Edilberto Strauss, Ph. D.

Examinador:

Flávio Luis Mello, D. Sc.

Examinador:

José Airton Chaves Cavalcante Junior, D. Sc.

Examinador:

Vinicius Drumond Gonzaga, M. Sc.

**Rio de Janeiro
Novembro de 2021**

Declaração de Autoria e de Direitos

Eu, **Caíque de Almeida Malafaia Souza** CPF 134.772.677-29, autor da monografia ***O USO DE FERRAMENTAS DE BI NO TRATAMENTO E VISUALIZAÇÃO DE DADOS EM UMA EMPRESA DE GRANDE PORTE***, subscrevo para os devidos fins, as seguintes informações:

1. O autor declara que o trabalho apresentado na defesa da monografia do curso de Pós-Graduação, Especialização MBA em Big Data, Business Intelligence e Business Analytics da Escola Politécnica da UFRJ é de sua autoria, sendo original em forma e conteúdo.
1. Excetuam-se do item 1 eventuais transcrições de texto, figuras, tabelas, conceitos e idéias, que identifiquem claramente a fonte original, explicitando as autorizações obtidas dos respectivos proprietários, quando necessárias.
1. O autor permite que a UFRJ, por um prazo indeterminado, efetue em qualquer mídia de divulgação, a publicação do trabalho acadêmico em sua totalidade, ou em parte. Essa autorização não envolve ônus de qualquer natureza à UFRJ, ou aos seus representantes.
2. O autor declara, ainda, ter a capacidade jurídica para a prática do presente ato, assim como ter conhecimento do teor da presente Declaração, estando ciente das sanções e punições legais, no que tange a cópia parcial, ou total, de obra intelectual, o que se configura como violação do direito autoral previsto no Código Penal Brasileiro no art.184 e art.299, bem como na Lei 9.610.
2. O autor é o único responsável pelo conteúdo apresentado nos trabalhos acadêmicos publicados, não cabendo à UFRJ, aos seus representantes, ou ao(s) orientador(es), qualquer responsabilização/ indenização nesse sentido.
1. Por ser verdade, firmo a presente declaração.

Rio de Janeiro, _____ de _____ de _____.

Caíque de Almeida Malafaia Souza

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

Av. Athos da Silveira, 149 - Centro de Tecnologia, Bloco H, sala - 212,
Cidade Universitária Rio de Janeiro – RJ - CEP 21949-900.

Este exemplar é de propriedade Escola Politécnica da Universidade Federal do Rio de Janeiro, que poderá incluí-lo em base de dados, armazenar em computador, microfilmear ou adotar qualquer forma de arquivamento.

Permitida a menção, reprodução parcial ou integral e a transmissão entre bibliotecas deste trabalho, sem modificação de seu texto, em qualquer meio que esteja ou venha a ser fixado, para pesquisa acadêmica, comentários e citações, desde que sem finalidade comercial e que seja feita a referência bibliográfica completa.

Os conceitos expressos neste trabalho são de responsabilidade do(s) autor(es).

AGRADECIMENTO

Agradeço a Deus por me dar a vida e a força que preciso para lutar por aquilo que acredito.

Agradeço aos meus pais e avós por investirem em minha educação e acreditarem em mim, me incentivando mesmo nos momentos difíceis da vida.

Agradeço a minha esposa por sempre estar ao meu lado, me dando forças e tornando os meus dias mais leves.

Agradeço aos meus professores que me acompanharam em minha jornada pelo conhecimento, em todas as etapas da minha vida.

RESUMO

Em tempos em que informação vale ouro, os dados são protagonistas dentro das organizações. Adotar boas práticas de armazenamento e tratamento de dados é premissa básica para uma empresa que busca crescer no mercado atual. Visando obter maior eficiência nas atividades de inteligência e extrair o máximo de informações a partir dos dados gerados, é grande o número de organizações que utilizam ferramentas de *Business Intelligence* em seus processos. Por ser um conceito que se relaciona à otimização da gestão dos recursos de informação das organizações, o *Business Intelligence* é um termo em alta no meio empresarial. O presente trabalho demonstra na prática o uso de algumas técnicas e ferramentas de *Business Intelligence* no tratamento de dados e geração de indicadores dentro da área de vendas de uma empresa de grande porte. Por meio de um caso de uso, são apresentadas as necessidades encontradas na área de vendas estudada e listadas as etapas presentes na arquitetura do processo de *Business Intelligence* aplicado na empresa. Foram desenvolvidas rotinas para os processos de extração, transformação e carga de dados e elaborados *dashboards* a partir das premissas definidas pela área de vendas apresentada no estudo de caso. Ao final do processo, foram percebidos ganhos em performance, tempo de processamento, qualidade da informação, padronização dos produtos de inteligência e mudança de hábito dos colaboradores, reforçando o impacto positivo da aplicação dos conceitos do *Business Intelligence* nas organizações.

Palavras-chave: Business Intelligence, ETL, Dashboards.

ABSTRACT

In times where information is gold, data are protagonists in companies. The adoption of good data storage and processing practices is a basic premise for a company seeking to grow in the current market. Aiming to obtain greater efficiency in intelligence activities and extract the maximum amount of information generated from the data, a large number of companies uses Business Intelligence tools in their processes. As a concept related to the optimization of companies' information resources management, Business Intelligence is a term on the rise in the business world. This work demonstrates in practice the use of some Business Intelligence techniques and tools in data processing and indicators generation in a large company sales area. Through a use case, the needs found in the studied sales area and the steps present in the Business Intelligence process architecture applied in the company are listed. Routines were developed for data extraction, transformation and loading processes and dashboards were prepared from the premises defined by the sales area presented in the study case. At the end of the process, gains were seen in performance, processing time, information quality, intelligence products standardization and change in employee habits, reinforcing the positive impact of Business Intelligence's concepts applying in companies.

Keywords: Business Intelligence, ETL, Dashboards.

SIGLAS

| | |
|---------------------|--|
| API | Interface de Programação Aplicada |
| BD | Banco de Dados |
| BI | Business Intelligence |
| DE | Diretor Executivo |
| DV | Diretor de Vendas |
| DW | Data Warehouse |
| EIS | Sistemas de Informações Executivas |
| FR1 | Fonte de Receita 1 |
| FR2 | Fonte de Receita 2 |
| FR3_anterior | Fonte de Receita 3 Ano Anterior |
| FR3_atual | Fonte de Receita 3 Ano Atual |
| GC | Gerente de Contas |
| GEV | Gerente Executivo |
| MIT | Massachusetts Institute of Technology |
| PDI | Pentaho Data Integration |
| SGBD | Sistema de Gerenciamento de Banco de Dados |
| SIG | Sistemas de Informações Gerenciais |
| SQL | Structured Query Language |
| TI | Tecnologia da Informação |

LISTA DE FIGURAS

| | | |
|-------------|--|----|
| Figura 2.1 | Evolução do BI | 4 |
| Figura 2.2 | Arquitetura de um processo de BI | 5 |
| Figura 2.3 | O SGBD faz o gerenciamento da interação entre o usuário final e o BD | 6 |
| Figura 2.4 | Relacionamento entre entidades num banco de dados relacional | 7 |
| Figura 2.5 | Exemplo de dado armazenado em um BD orientado a documentos | 8 |
| Figura 2.6 | Exemplo de dado armazenado em um BD colunar | 9 |
| Figura 2.7 | Exemplo de dado armazenado em um BD orientado a grafos | 9 |
| Figura 2.8 | Exemplo de dado armazenado em um BD do tipo chave-valor | 10 |
| Figura 2.9 | Exemplo de dashboard vendas no Tableau | 13 |
| Figura 2.10 | Exemplos de elementos de visualização de dados | 14 |
| Figura 3.1 | Exemplificação do organograma da área de vendas | 16 |
| Figura 3.2 | Arquitetura de BI proposta | 17 |
| Figura 3.3 | Exemplo de um fluxo de ETL desenvolvido no PDI | 22 |
| Figura 3.4 | Fluxo de transformação 1: ajuste e concatenação das bases de receita | 23 |
| Figura 3.5 | Fluxo de transformação 2: enriquecimento da base de clientes | 24 |
| Figura 3.6 | Fluxo de transformação 3: enriquecimento da base FR_Total | 25 |

| | | |
|-------------|---|----|
| Figura 3.7 | Tela inicial do Cortex Connector | 27 |
| Figura 3.8 | Tela de configuração de uma data source tipo Projeto Kettle | 27 |
| Figura 3.9 | Configuração de uma data connection: selecionando uma data source | 28 |
| Figura 3.10 | Configuração de uma data connection: selecionando o data set | 28 |
| Figura 3.11 | Configuração de uma data connection: agendando a execução | 29 |
| Figura 3.12 | Plataforma Cortex: interface para elaboração de análises | 31 |
| Figura 3.13 | Plataforma Cortex: interface para elaboração de dashboards | 31 |
| Figura 4.1 | Dashboard desenvolvido: Receita por diretorias | 34 |
| Figura 4.2 | Dashboard desenvolvido: Receita por plataformas | 35 |
| Figura 4.3 | Dashboard desenvolvido: Receita por famílias | 36 |
| Figura 4.4 | Dashboard desenvolvido: Receita por serviços | 37 |
| Figura 4.5 | Dashboard desenvolvido: Receita por verticais | 39 |
| Figura 4.6 | Dashboard desenvolvido: Receita por clientes | 40 |
| Figura 4.7 | Alguns dashboards mobile desenvolvidos | 41 |

LISTA DE QUADROS

| | | |
|------------|---|----|
| Quadro 3.1 | Estrutura e descrição das bases de dados FR1 e FR2 | 18 |
| Quadro 3.2 | Estrutura e descrição das bases de dados FR3_atual e FR3_anterior | 19 |
| Quadro 3.3 | Estrutura e descrição da base de dados FR4 | 19 |
| Quadro 3.4 | Estrutura e descrição da base de dados Clientes | 20 |
| Quadro 3.5 | Estrutura e descrição da base de dados da Estrutura Comercial | 21 |
| Quadro 3.6 | Estrutura da base de serviços por famílias e plataformas | 25 |
| Quadro 3.7 | Base FR_Final: estrutura da versão final gerada no tratamento dos dados | 26 |
| Quadro 3.8 | Exemplo da composição do campo ‘CONTROLE_ACESSO’ | 30 |

Sumário

| | |
|--|-----------|
| Capítulo 1: Introdução | 1 |
| 1.1 – Tema | 1 |
| 1.2 – Justificativa | 1 |
| 1.3 – Objetivos | 1 |
| 1.4 – Delimitação | 2 |
| 1.5 – Metodologia | 2 |
| Capítulo 2: Embasamento Teórico | 3 |
| 2.1 – Business Intelligence | 3 |
| 2.1.1 – Evolução | 3 |
| 2.1.2 – Conceituação | 3 |
| 2.1.3 – Arquitetura | 5 |
| 2.2 – Fontes de Dados | 5 |
| 2.2.1 – Banco de Dados Relacional | 6 |
| 2.2.2 – Banco de Dados não Relacional | 7 |
| 2.3 – ETL | 10 |
| 2.4 – Data Warehouse | 11 |
| 2.5 – Apresentação | 12 |
| Capítulo 3: Propostas Tecnológicas | 15 |
| 3.1 – Estrutura da Área. | 15 |
| 3.2 – Processo de Geração da Informação pré BI. | 16 |
| 3.3 – Necessidades Encontradas no Processo de Geração da Informação. | 17 |
| 3.4 – Solução de BI Proposta. | 17 |
| 3.4.1 – Fontes de Dados | 18 |
| 3.4.2 – Extração dos Dados. | 21 |
| 3.4.3 – Transformação dos Dados. | 21 |
| 3.4.4 – Carga dos Dados. | 26 |
| 3.4.5 – Armazenamento dos Dados. | 29 |
| 3.4.6 – Configuração de Acesso aos Registros. | 29 |
| 3.4.7 – Desenvolvimento dos Dashboards. | 30 |
| Capítulo 4: Resultados Obtidos | 33 |
| 4.1 – Dashboards Gerados no Processo. | 33 |
| 4.1.1 – Receita por Diretorias. | 34 |
| 4.1.2 – Receita por Plataformas. | 35 |
| 4.1.3 – Receita por Famílias. | 36 |
| 4.1.4 – Receita por Serviços. | 37 |
| 4.1.5 – Receita por Verticais. | 38 |
| 4.1.6 – Receita por Clientes. | 39 |
| 4.1.7 – Dashboards Mobile. | 40 |
| 4.2 – Comparação de Cenários. | 40 |

| | |
|--|-----------|
| 4.3 – Ganhos Obtidos. | 42 |
| Capítulo 5: Conclusão e Trabalhos Futuros | 43 |
| 5.1 – Conclusão | 43 |
| 5.2 – Trabalhos Futuros | 44 |
| Referências Bibliográficas | 45 |

Capítulo 1

Introdução

1.1 – Tema

Encontrar a melhor rota para um destino desconhecido nunca foi tão fácil: basta inserir o endereço no aplicativo do seu smartphone e iniciar o trajeto. Pedir comida em casa, comprar pela internet, pagar contas via *QR Code* ou estudar à distância são exemplos de atividades que foram incorporadas ao nosso cotidiano trazendo benefícios que mudaram radicalmente nossos hábitos, formas de consumir, comunicar e até mesmo enxergar o mundo. Tais mudanças presentes nos indivíduos são também notórias nas organizações, que passaram a adotar ferramentas e técnicas para melhor gerir seus processos e informações. Entre elas estão a automatização de rotinas de dados, padronização de tabelas, centralização de sistemas e elaboração de algoritmos de aprendizado de máquina.

Neste contexto marcado por constantes mudanças e evoluções nas empresas, encontramos o protagonismo da inteligência de negócios, muito conhecida por seu termo em inglês *business intelligence* (BI), que abarca um vasto conjunto de práticas e ferramentas com foco em dados, sendo vital para a sobrevivência e evolução de qualquer empresa que almeja sucesso no mercado atual.

1.2 – Justificativa

O correto uso do BI dentro das organizações impacta de forma direta o processo de geração de indicadores, produzindo ganhos expressivos como a padronização das bases de dados, a otimização das consultas e a facilidade de acesso às informações. Além dos ganhos operacionais, são percebidos também ganhos gerenciais, como um melhor entendimento do perfil do cliente, identificação de novas oportunidades de negócios, redução de erros e custos e a tomada de decisões baseadas em dados. Considerando o impacto positivo do uso do BI e seu poder de otimizar processos como o tratamento e visualização de dados, além da capacidade de modificar hábitos nas organizações, o presente trabalho busca explorar a implantação e utilização de ferramentas de BI em uma grande empresa do mercado de telecomunicações.

1.3 – Objetivos

O objetivo principal deste trabalho é, apresentar na prática, o desenvolvimento do processo de estruturação de uma plataforma de inteligência de negócios dentro de uma empresa real, passando pelas etapas presentes na arquitetura básica de um projeto de BI, indicando as técnicas e processos utilizados e apresentando os *dashboards* gerados no processo. Ao final, diante da comparação entre os cenários anterior e atual da área referida, busca-se mensurar alguns dos ganhos percebidos após a implementação do processo e o uso das ferramentas de BI.

1.4 – Delimitação

O trabalho apresentado a seguir foi desenvolvido em uma empresa de grande porte do mercado de telecomunicações brasileiro, possuindo uma carteira de mais de 47 mil clientes. Os processos apresentados neste trabalho foram realizados na área de vendas da empresa em questão, área esta que possui 580 colaboradores espalhados em todo o território nacional.

Vale ressaltar que os dados exibidos neste trabalho são fictícios, devido a estes possuírem condições sensíveis e confidenciais, sendo necessária à sua substituição. Em relação às ferramentas de BI, foram utilizadas no processo aquelas disponibilizadas e contratadas pela referida empresa no momento da implementação, o que não exclui ou inviabiliza outras práticas e ferramentas, assim como não torna a opção apresentada neste trabalho como a única ou a melhor possível.

1.5 – Metodologia

A metodologia utilizada foi um estudo de caso, que consiste em uma profunda análise de um ou mais objetos possibilitando um conhecimento detalhado a cerca dos mesmos (GIL, 1996; BERTO; NAKANO, 2000). Esta metodologia visa esclarecer o motivo pelo qual uma decisão ou um conjunto de decisões foram tomadas, a forma de sua implementação e quais foram os resultados alcançados (YIN, 2001).

Neste contexto, foram mapeadas as necessidades da área objeto do estudo, traçadas as diretrizes para a implementação do projeto, desenvolvida a proposta tecnológica e mensurados os ganhos ao final do processo.

Capítulo 2

Embasamento Teórico

2.1 – Business Intelligence

2.1.1 – Evolução

As raízes do BI estão na década de 1970, quando seus conceitos, ainda que mais rudimentares, já eram utilizados em Sistemas de Informações Gerenciais (SIG). Entre o fim da década de 70 e início dos anos 1980, o termo Sistemas de Informações Executivas (EIS, do inglês *Executive Information Systems*) veio à tona com base em trabalhos desenvolvidos pelo Instituto de Tecnologia de Massachusetts (MIT, do inglês *Massachusetts Institute of Technology*) (FURLAN; IVO; AMARAL, 1994). Segundo Pozzebon e Freitas (1996, p. 29) “EIS é uma solução em termos de informática que disponibiliza informações corporativas e estratégicas para os decisores de uma organização, de forma a otimizar sua habilidade para tomar decisões de negócios importantes”. Somente nos anos 1990 que o termo BI foi adotado pelo *Gartner Group*, sendo caracterizado posteriormente como uma resultante das transformações ocorridas no modelo dos EIS, tais como o aumento do poderio de capacidade analítica e a adição de funcionalidades de inteligência artificial (SHARDA; DELEN; TURBAN, 2013).

2.1.2 – Conceituação

“BI tem o potencial de afetar positivamente a cultura de uma empresa criando a ‘sabedoria do negócio’ e distribuindo-a a todos os usuários em uma organização” (CORONEL; MORRIS, 2018, p. 515).

Por se tratar de um termo originado a partir de conceitos diversos e por ser resultante de transformações e agregações ao longo do tempo, o BI recebe uma definição mais abrangente, sendo caracterizado não somente como uma ferramenta ou técnica específica, mas por um conjunto de práticas e conceitos. Essa visão é reforçada por Sharda (2013, p. 14) ao dizer que "*Business Intelligence* (BI) é um termo guarda-chuva que combina arquiteturas, bases de dados, ferramentas analíticas, aplicações e metodologias”.

Dessa forma, o BI pode ser designado para retratar recursos de informação de uma organização visando desenvolver uma compreensão mais precisa das dinâmicas do negócio. Semelhantemente, torna-se possível a tomada de melhores decisões por meio da coleta de informações de fontes diversas voltadas para a definição de estratégias competitivas nos negócios da empresa (ARULDOSS; LAKSHMI TRAVIS; PRASANNA VENKATESAN, 2014; LI; SHUE; LEE, 2008). Além disso, os conceitos de estruturas de dados como bancos de dados tradicionais, *data warehouse* e *data marts*, podem ser incorporados à definição de BI visando o tratamento relacional dessas informações, assim como as técnicas de *data mining* aplicadas sobre elas (BARBIERI, 2011). A representação do grande conjunto de técnicas e ferramentas que podem ser inseridas no escopo do BI está retratada na figura 2.1.

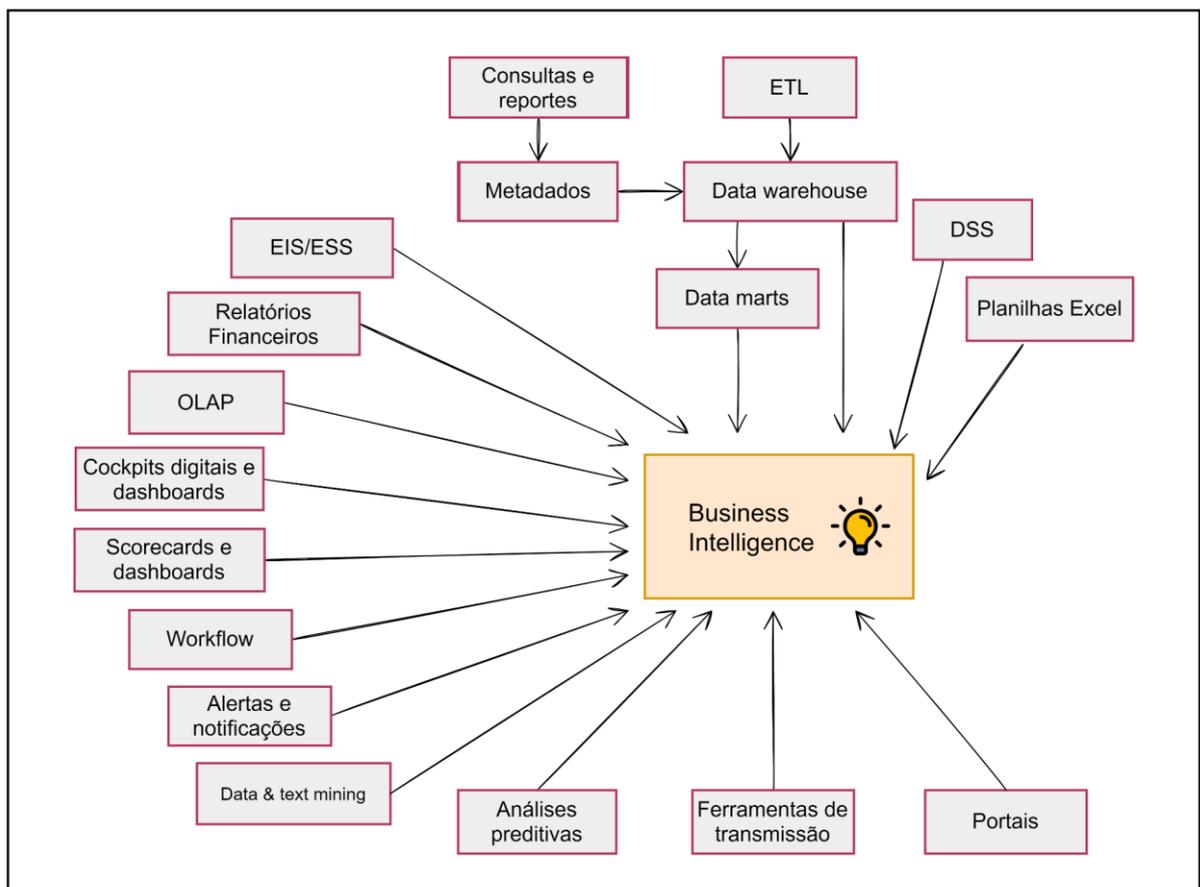


Figura 2.1 – Evolução do BI.
 Fonte: Modificado de SHARDA; DELEN; TURBAN, 2017.

Como exemplos de algumas aplicações e áreas do BI tem-se: i) análise de dados (*Data Mining* e OLAP); ii) ferramentas de ETL (*Talend* e o *Pentaho data integration*); iii) fontes de dados (portais, planilhas excel e *data warehouse*); iv) produtos de inteligência (*Cockpits, dashboards, alertas e notificações*).

2.1.3 – Arquitetura

A arquitetura básica do processo de BI consiste em um pipeline de integração de dados (Figura 2.2) que se inicia no mapeamento das fontes de dados. Em seguida, ocorre a extração, tratamento e carga dos dados em um *Data Warehouse* (DW), etapa conhecida como ETL (do inglês *Extract Transform Load*). O DW, então, armazena e disponibiliza os dados para análise e desenvolvimento de produtos de inteligência, que são compartilhados e consumidos pelos usuários (DAYAL et al., 2009). Estas etapas serão detalhadas nos tópicos a seguir.

É importante ressaltar que são diversas as possibilidades e componentes que podem ser inseridos e parametrizados dentro da arquitetura de um projeto de BI. Tais componentes são diretamente influenciados pela demanda a ser desenvolvida, regras de negócio, tecnologias disponíveis e nível de conhecimento dos profissionais que o desenvolvem. Dessa forma, não há necessariamente um único modelo para cada demanda.

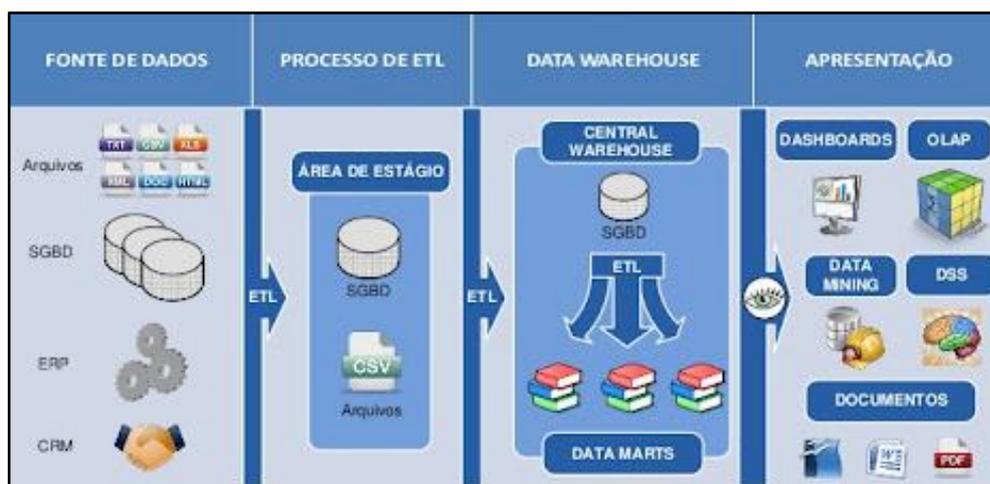


Figura 2.2 - Arquitetura de um processo de BI.
Fonte: LORENA, 2011, adaptado de BOUMAN; DONGEN, 2009.

2.2 – Fontes de Dados

O entendimento da necessidade do cliente é o norteador de todo o desenvolvimento de um projeto. É a partir desse entendimento que são definidas as fontes de dados que serão utilizadas.

Numa visão geral podemos dividir as fontes de dados entre internas e externas. Pelo ponto de vista de uma organização, as fontes externas como o próprio nome diz, têm origem externa à organização. Alguns exemplos são os dados do censo, variação do preço de ações, estatísticas eleitorais, ou até mesmo informações de redes sociais, geralmente obtidas através de uma Interface de Programação Aplicada (API), como Facebook ou Twitter. Quando se fala

das fontes de dados internas, refere-se àquelas que a própria organização cria, controla ou coleta, como por exemplo os dados de venda da companhia ou informações de RH. Os dados internos geralmente estão dispostos em planilhas Excel, arquivos de texto, sistemas ou banco de dados (BD).

De forma mais detalhada, os BDs consistem em uma estrutura computacional compartilhada e integrada que armazena um conjunto de dados e/ou metadados (dados sobre dados). Um BD parece um arquivo eletrônico organizado onde um software conhecido como Sistema de Gerenciamento de Banco de Dados (SGBD) gerencia sua interface e acessos aos dados. Em outras palavras, o SGBD funciona como intermediário entre o BD e o usuário, recebendo as solicitações das aplicações e as traduzindo em operações complexas para atendê-las. O usuário, por sua vez, experimenta uma interface muito mais amigável, visto que boa parte da complexidade interna do BD é ocultada pelo SGBD (Figura 2.3) (CORONEL; MORRIS, 2018). Como exemplos de SGBDs, podemos citar: Oracle, MySQL, PostgreSQL e MongoDB.

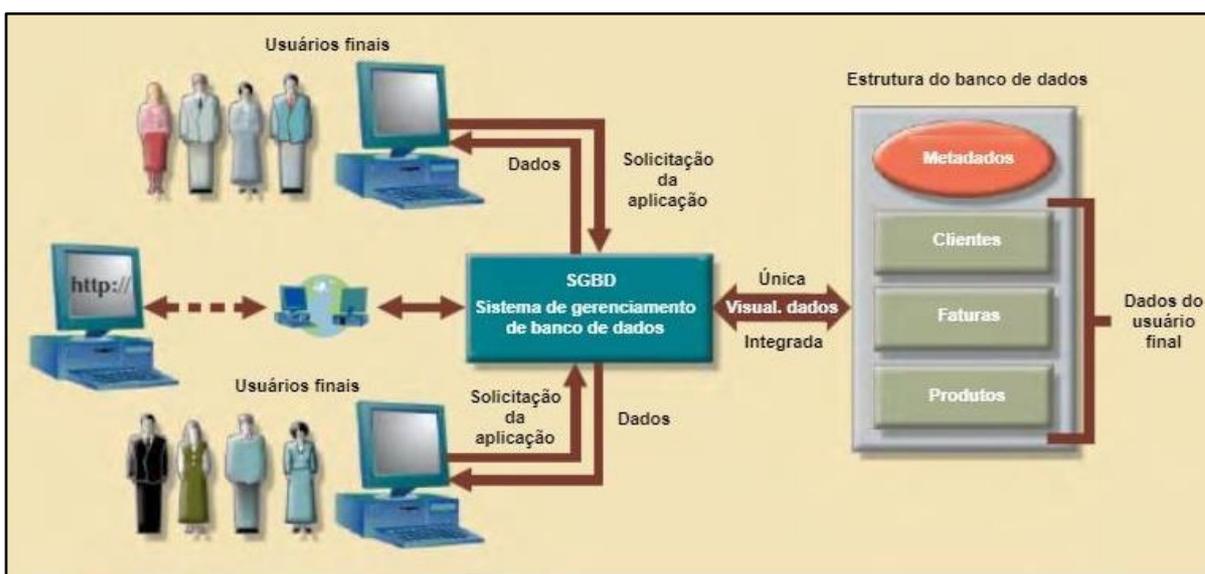


Figura 2.3 - O SGBD faz o gerenciamento da interação entre o usuário final e o BD.
Fonte: traduzido de CORONEL; MORRIS, 2018.

Existem basicamente duas categorias de BD, a relacional e a não relacional, que serão mais bem exploradas nos próximos tópicos. A escolha da mais apropriada a uma empresa está totalmente ligada ao tipo de regra do negócio aplicada.

2.2.1 – Banco de dados relacional

Um BD relacional é aquele onde o armazenamento dos dados é feito em um formato estruturado, usando linhas e colunas. Este modelo é baseado no sistema de armazenamento das

informações em tabelas, conhecidas como entidades (Figura 2.4), onde cada linha de informação é composta por valores definidos em suas colunas (AGUENA, 2018). Essa estrutura facilita o acesso e localização de valores pontuais no BD, tornando possível a execução de consultas em várias tabelas de forma simultânea (TECHLIB, 2005).

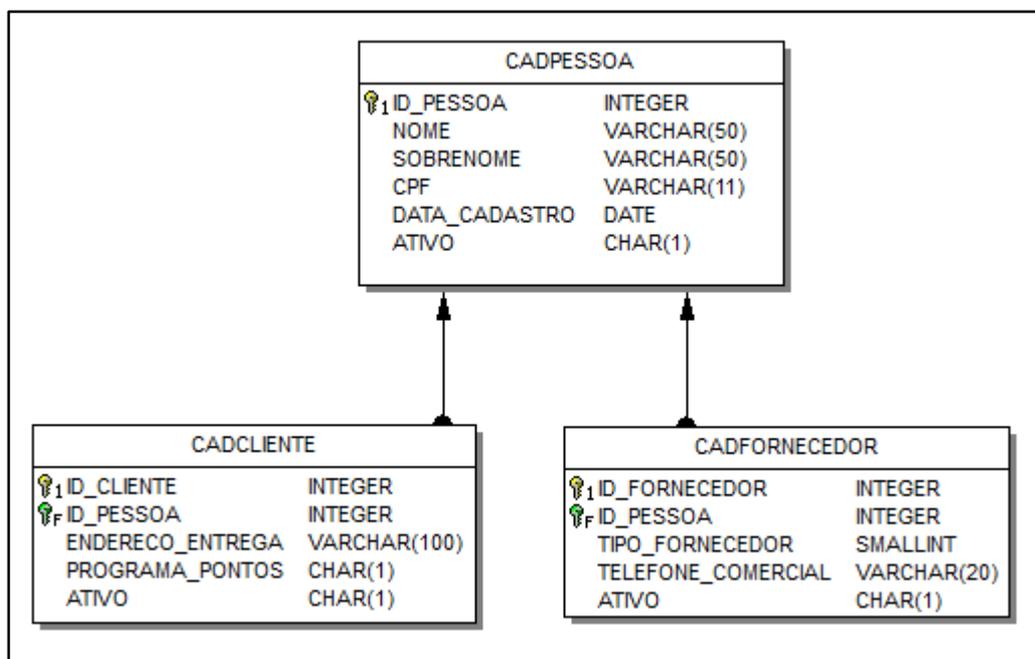


Figura 2.4 - Relacionamento entre entidades num banco de dados relacional.
Fonte: AGUENA, 2018.

Dentre as características do BD relacional, destaca-se sua linguagem de consulta, que é baseada no *Structured Query Language* (SQL), ou “Linguagem de Consulta Estruturada” traduzindo para o português. Além disso, um BD relacional tem como base os pilares de Atomicidade, Consistência, Isolamento e Durabilidade, representados pela sigla ACID (NOLETO, 2020). Estes pilares têm como finalidade a redução de possíveis anomalias e proteção da integridade do BD, definindo exatamente como as transações ocorrem. Diferentemente dos BDs relacionais, os BDs não relacionais têm um objetivo principal de flexibilidade e velocidade (FERNANDES, 2020).

2.2.2 – Banco de dados não relacional

Um BD não relacional é também conhecido como BD NoSQL, isto porque não usam como linguagem de consulta o SQL, mas sim outras linguagens de programação. Além disso, o BD não relacional usa um modelo otimizado para o armazenamento de dados, diferente do BD relacional, que utiliza o sistema de tabelas (TEJADA, 2018). Dessa forma, observa-se uma

escalabilidade mais horizontal e foco em melhor performance, visando suprir as necessidades encontradas nos BDs relacionais (RAFAEL, 2019). Os principais tipos de BDs NoSQL são:

- Orientado a documentos (Figura 2.5) - os dados são armazenados como documentos, onde o valor de cada campo pode ser um item escalar (um número, um elemento composto, uma lista). Os documentos podem ser, por exemplo, descritos como dados no padrão JSON (RAFAEL, 2019). Esse tipo de repositório não demanda a mesma estrutura em todos os documentos, o que confere grande flexibilidade (TEJADA, 2018). Um exemplo de banco de dados neste formato é o MongoDB.

| Key | Document |
|------|---|
| 1001 | <pre>{ "CustomerID": 99, "OrderItems": [{ "ProductID": 2010, "Quantity": 2, "Cost": 520 }, { "ProductID": 4365, "Quantity": 1, "Cost": 18 }], "OrderDate": "04/01/2017" }</pre> |
| 1002 | <pre>{ "CustomerID": 220, "OrderItems": [{ "ProductID": 1285, "Quantity": 1, "Cost": 120 }], "OrderDate": "05/08/2017" }</pre> |

Figura 2.5 - Exemplo de dado armazenado em um BD orientado a documentos.
Fonte: TEJADA, 2018.

- Colunar (Figura 2.6) - o armazenamento dos dados é feito em colunas e linhas, suportando também sub-colunas e sub-linhas (RAFAEL, 2019). Assim, colunas de uma família de colunas são armazenadas juntas dentro de um mesmo arquivo que possui um número determinado de linhas. Esse método reduz a quantidade de dados a serem lidos do disco, o que gera uma melhora significativa de desempenho em grandes conjuntos de dados (TEJADA, 2018). Um exemplo de banco de dados colunar é o Cassandra.

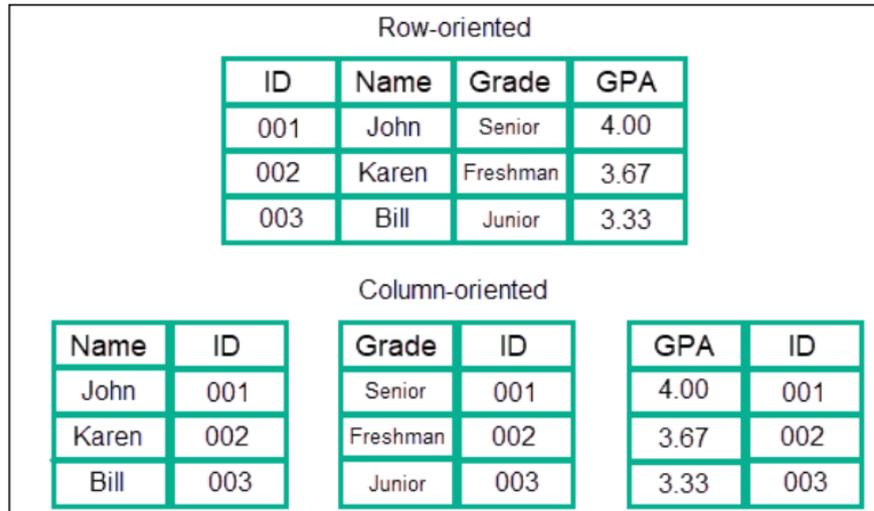


Figura 2.6 - Exemplo de dado armazenado em um BD colunar.
 Fonte: WILLIAMS, 2021.

- Orientado a grafos (Figura 2.7) - nesse modelo, os dados são armazenados na forma de grafos (vértices e arestas) visando garantir eficácia em suas consultas. Os vértices representam entidades e as arestas delimitam as relações entre as entidades, podendo ter uma direção indicando a natureza do relacionamento (RAFAEL, 2019; TEJADA, 2018). O Neo4j é um BD NoSQL que utiliza grafos.

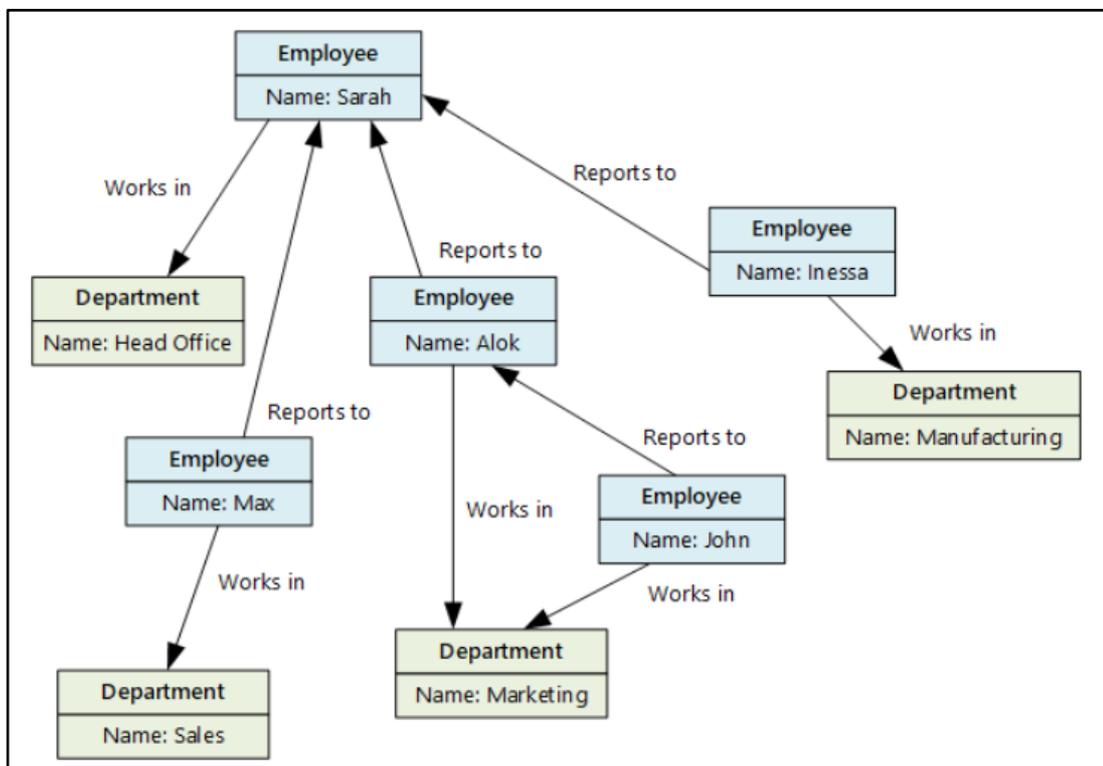


Figura 2.7 - Exemplo de dado armazenado em um BD orientado a grafos.
 Fonte: TEJADA, 2018.

- Tipo chave-valor (Figura 2.8) - essa categoria de BD NoSQL aguenta grandes cargas de dados. O conceito aqui é que cada valor de dados seja associado a uma chave única, apresentando uma combinação de chave-valor (RAFAEL, 2019). Um exemplo de BD NoSQL de chave-valor é Riak.

| Key | Value |
|--------|------------------------------|
| AAAAA | 1101001111010100110101111... |
| AABAB | 1001100001011001101011110... |
| DFA766 | 0000000000101010110101010... |
| FABCC4 | 1110110110101010100101101... |

Figura 2.8 - Exemplo de dado armazenado em um BD do tipo chave-valor.
Fonte: TEJADA, 2018.

2.3 – ETL

O processo de ETL tem sido utilizado há tempos por empresas como meio para alcançar uma visão consolidada de seus dados. Atualmente, tal método ainda é um componente vital no conjunto de ferramentas de integração de dados de uma organização (SAS, [s.d.]).

Como mencionado anteriormente, a sigla ETL vem do inglês “*Extract, Transform and Load*”. Em outras palavras, o ETL consiste em um processo de integração de dados que se divide em três etapas (extração, transformação e carregamento) e é utilizado na combinação de dados geralmente de fontes diversas (SAS, [s.d.]). Dentre as ferramentas de ETL disponíveis no mercado, podemos citar o *IBM Information Server (DataStage)*, o *Oracle Data Integrator (ODI)* e o *Microsoft Integration Services (SSIS)*. Pode-se encontrar também conjuntos de ferramentas de código aberto (*open source*), como o *Pentaho Data Integration (PDI)* e *Talend ETL* (CETAX, 2020).

A etapa de extração, primeira do processo de ETL, consiste, segundo o próprio nome, na coleta e extração de dados brutos de uma ou diversas fontes para posterior integração em um repositório. Dessa forma, após o mapeamento das fontes de dados de interesse, os mesmos devem ser coletados e armazenados em uma área intermediária provisória por meio de rotinas de extração (PEREIRA, 2000).

A segunda etapa do processo de ETL consiste na transformação desses dados brutos, com base nas regras estabelecidas, em dados com a formatação necessária para uso nas aplicações que virão a seguir. Isso significa que o dado deverá ser mapeado, limpo e tratado

para atender as necessidades operacionais (DATABRICKS, [s.d.]). A fase de transformação é útil na correção de imperfeições nas bases de dados originais, trazendo qualidade e coerência nas informações exibidas aos usuários finais do DW. Abaixo estão listados alguns dos problemas que podem ser encontrados e corrigidos durante o processo (FREITAS, 2002):

- Diferenças de unidade: informações que são complementares e exibem dados em diferentes unidades. Exemplo: prazo descrito em ‘anos’ em um item e em ‘dias’ em outro;
- Diferenças de precisão: valores apresentados com casas decimais diferentes;
- Diferenças de códigos ou expressões: códigos iguais representando informações distintas Exemplo: ‘SP’ significa São Paulo em um item e São Patrício em outro;
- Diferenças de granularidade: um campo totaliza as vendas em uma semana e outro em um mês;
- Diferenças de abstração: endereço composto por campos separados com rua, número e bairro em uma tabela, porém em outra o endereço está todo em um único campo.

Por fim, a etapa de carga consiste no processo de escrita dos dados tratados anteriormente em um repositório de dados, que geralmente é um DW. Dependendo dos requisitos da aplicação, o processo de carga pode ser bem simples ou nem tanto (DATABRICKS, [s.d.]).

2.4 – Data Warehouse

Grande parte da filosofia que rege os princípios básicos da tecnologia do DW foi expressa por Inmon, considerado o ‘pai’ do DW, em 1994. Ele o definiu como um depósito de dados orientado ao assunto, não volátil, integrado, variante no tempo, para apoiar a tomada de decisão (INMON; KELLEY, 1994). De forma prática, DW é um BD constantemente alimentado com dados de fontes externas e/ou sistemas operacionais, que são selecionados, depurados e organizados de maneira a agilizar o processamento de consultas, buscando criar uma visão única e real da empresa, assim como uma visão mais precisa do conjunto de dados. Em resumo, é um BD, geralmente de leitura, voltado para a análise e processamento de dados (CAMPOS; ROCHA, 1997; CORONEL; MORRIS, 2018).

Além de conceituarem, Inmon e Kelley criaram 12 regras que definem um DW, fornecendo uma visão precisa de seu escopo e propriedades (INMON; KELLEY, 1994). São elas:

1. O DW e os sistemas operacionais são separados;
2. Os dados do DW são integrados;
3. O DW contém dados históricos ao longo do tempo;
4. Os dados do DW são uma foto dos dados capturados em um certo ponto no tempo;
5. Os dados do DW são orientados ao sujeito;
6. Os dados DW são, em sua maioria, apenas de leitura e com lotes de atualizações periódicas de dados operacionais (não são permitidas atualizações online);
7. O desenvolvimento do ciclo de vida do DW, que é baseado em dados, difere do desenvolvimento dos sistemas clássicos, que é baseado em processos;
8. O DW contém dados com vários níveis de detalhes: de dados atuais, de dados antigos, pouco sumarizados e muito sumarizados;
9. O ambiente do DW é caracterizado por transações apenas de leitura para uma grande quantidade de dados. Já o ambiente operacional é caracterizado por várias transações de atualização para uma pouca quantidade de entidades de dados por vez;
10. O DW possui um ambiente que mapeia as fontes de dados, transformações e armazenamento;
11. Os metadados do DW são um componente crítico deste ambiente, pois são eles que identificam e definem todos os elementos de dados. Através dos metadados que se conhece a fonte, transformação, integração, armazenamento, uso, relacionamentos e história de cada elemento;
12. O DW contém um mecanismo de reversão para o uso dos recursos que impõe a otimização do uso dos dados pelos usuários finais.

2.5 – Apresentação

Todo processo que vai desde o mapeamento das fontes de dados até ao carregamento dos registros tratados no DW culmina em insumos que são utilizados na elaboração dos produtos de inteligência. Estes produtos são consumidos por diretores, gerentes, analistas, ou qualquer indivíduo envolvido na tomada de decisão. A elaboração dos mesmos é geralmente realizada através de ferramentas de BI que possuem, em seu pacote de funcionalidades, recursos de visualização de dados para a produção de análises e indicadores, que irão compor *dashboards* (Figura 2.9) e alertas auxiliando na compreensão dos dados. Dessa forma, a etapa de apresentação dos dados se torna tão importante quanto as anteriores, afinal, toda e qualquer

informação gerada e devidamente interpretada é capaz de se tornar conhecimento convertido em ações visando o crescimento da empresa.

Com o passar do tempo, tem se criado uma grande necessidade pela agilidade do processo decisório dentro das empresas. Isso faz com que, cada vez mais, o uso de *dashboards* e painéis gerenciais de negócio estejam presentes nessa etapa, compilando vários indicadores em pouco espaço de tela e, assim, tornando a leitura dos dados fácil e rápida. São exemplos de ferramentas de BI o Tableau, QlikView, Power BI e a Plataforma Cortex.



Figura 2.9 - Exemplo de *dashboard* vendas no Tableau.
Fonte: INTELLIGENCE PARTNER, 2015.

A visualização de dados refere-se ao estudo e criação da representação dos dados de maneira visual, como diagramas, gráficos e mapas (Figura 2.10). Seu objetivo é fazer um link entre o operacional e o executivo, com a finalidade de ver e entender exceções, tendências e padrões nos dados, possibilitando a rápida tomada de decisão através de análises e relatórios (PINO, 2016; TABLEAU, [s.d.]).

O campo da visualização de dados combina arte e ciência, já que as análises, além de criativas e visualmente agradáveis, precisam ser assertivas, úteis e funcionais para o usuário final (NEDIGER, 2020). Assim, uma boa visualização de dados deve contar uma história, compilando os dados de maneira mais compreensível, eliminando elementos desnecessários e ressaltando informações úteis (TABLEAU, [s.d.]).

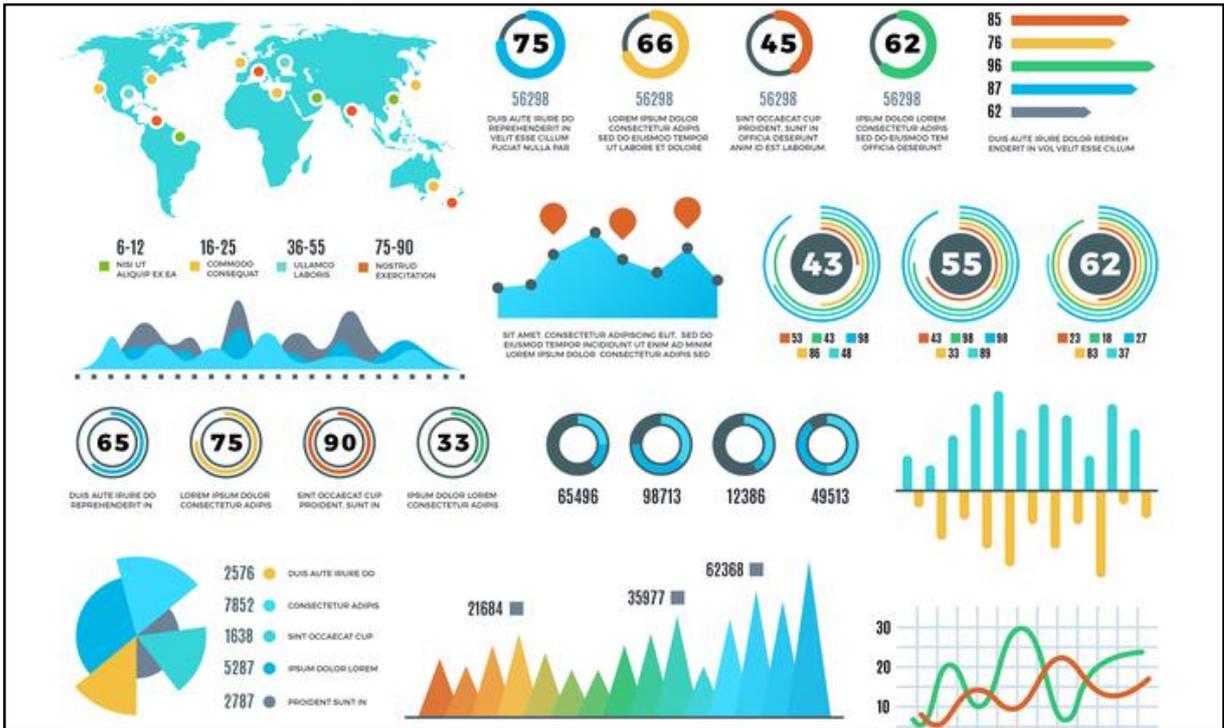


Figura 2.10 - Exemplos de elementos de visualização de dados.
 Fonte: PATEL, [s.d.].

Alguns benefícios da visualização de dados são (FIVEACTS, 2020):

- Ajudar a identificar padrões: o desenvolvimento de técnicas de visualização de dados possibilitou identificar padrões de comportamento e tendências que antes eram apenas números empilhados;
- Favorecer a interação: softwares com conceito de visualização de dados facilitam a assimilação de informações, visto que a disposição dos elementos em uma interface simples torna mais fácil a interpretação e a comunicação;
- Facilitar o tratamento da informação: o tratamento da informação se torna muito mais simples quando os dados são dispostos em diagramas e gráficos, facilitando seu entendimento sem necessidade de tradução;
- Permite descobrir correlações: ferramentas como o *Google Analytics*, projetadas com princípios de visualização de dados, permitem o cruzamento de dados, possibilitando a descoberta de correlações que poderiam não ser notadas.

Capítulo 3

Propostas Tecnológicas

A solução proposta neste trabalho e seus desdobramentos têm como objetivo a criação de uma plataforma de BI para a área de vendas que será apresentada a seguir. A ideia é possibilitar aos colaboradores o consumo dos indicadores e resultados de vendas através de dashboards dinâmicos e intuitivos, gerados a partir de um processo de ETL 100% automatizado. O produto final deve ainda conter uma rotina de notificação automática, estar armazenado na nuvem e poder ser acessado por dispositivos móveis.

3.1 – Estrutura da Área

O estudo de caso apresentado neste trabalho foi realizado na área de vendas de uma empresa de grande porte do setor de telecomunicações, já consolidada no mercado brasileiro e com atuação em todo o território nacional. A área de vendas em questão lida com um portfólio de mais de 60 serviços distribuídos em uma carteira de mais de 47 mil clientes. Sua estrutura é constituída de 580 colaboradores distribuídos em (Figura 3.1):

- Presidente da empresa - participa das grandes negociações e recebe informações dos seus subordinados diretos;
- Diretores executivos (DEs) - lideram 5 diretorias executivas divididas por perfil de clientes (tamanho, área de atuação, localidade e receita) e respondem ao presidente;
- Diretores de vendas (DVs) - estão abaixo dos DEs e possuem sua porção de clientes dentro da carteira dos seus superiores;
- Gerentes executivos (GEVs) - têm suas carteiras de cliente constituídas como parte das carteiras dos DVs e respondem diretamente a estes;
- Gerentes de contas (GCs) - são subordinados diretos dos GEVs e estão na base do organograma da área de vendas. Em alguns casos, os GCs possuem apenas um cliente em suas carteiras ou até mesmo dividem clientes;
- Suporte comercial - equipe de apoio às vendas presentes em cada DE, funcionando como um suporte nas atividades dentro da diretoria.

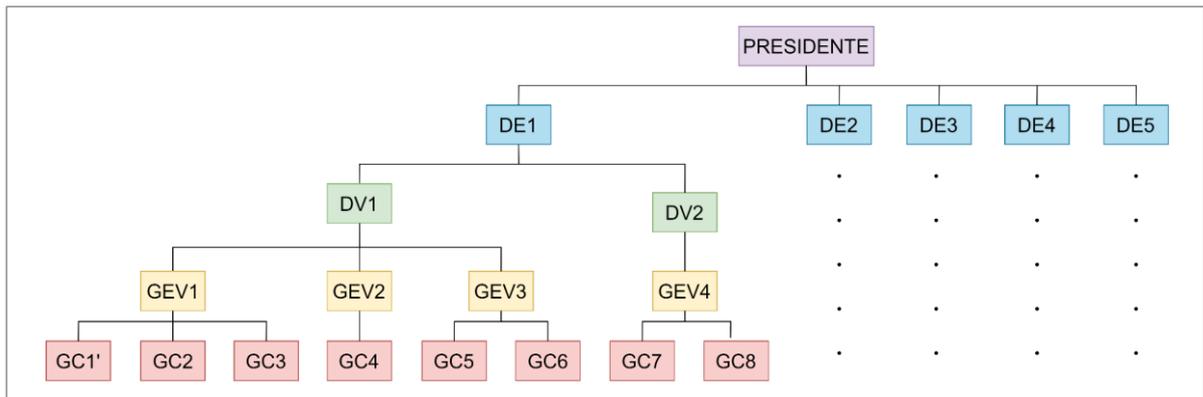


Figura 3.1 - Exemplificação do organograma da área de vendas.
 Fonte: Elaborado pelo Autor.

A visão dos dados e informações no processo de inteligência da área de vendas também obedece a uma hierarquia: o presidente da empresa visualiza o resultado de todas as diretorias executivas e seus subníveis, porém a partir dos DEs, a visão passa a ser limitada pela diretoria executiva em que está inserido o colaborador e sua posição dentro do organograma. Dessa forma, diferentemente do presidente, um DE visualiza toda a informação de sua diretoria executiva, mas não tem visibilidade dos dados de outras diretorias. Seguindo essa linha de raciocínio, um DV só visualiza seus dados e os de seus subordinados diretos. Esse racional se repete para GEVs e GCs, sendo estes limitados somente aos próprios dados, visto que não possuem subordinados diretos.

3.2 – Processo de geração da informação pré-BI

Para a compilação dos dados e geração das informações mensais de vendas, era realizado todos os meses um processo manual e exaustivo. Os dados de vendas oriundos de cinco bases de dados (fontes de receita distintas dentro da empresa) eram armazenados em um BD Oracle, gerido por uma equipe específica. Essa equipe distribuía os dados de cada diretoria executiva para seus respectivos suportes comerciais, e estes redistribuíam as informações para os níveis específicos das diretorias. Uma outra área, responsável pela gestão da informação dos clientes, disponibilizava uma base com a carteira de clientes de cada diretoria, que também era repassada para os demais colaboradores. Com os dados de clientes e receita (vendas) em mãos, cada colaborador trabalhava suas informações via excel, validando os dados e construindo seus próprios indicadores para controle interno.

Para a reunião mensal de vendas com o presidente, era elaborada uma grande apresentação no powerpoint, agregando os resultados obtidos por todas diretorias executivas (soma dos resultados de todos os colaboradores de sua estrutura). As áreas de suporte comercial

trabalhavam resumindo as informações das diretorias e elaborando os indicadores para a apresentação dos resultados num processo contínuo. Quando o material de uma reunião estava pronto, o suporte já iniciava os preparativos do processo de elaboração do conteúdo para a próxima reunião. Por haver muitas etapas e interferências no processo, era normal que fossem levantados valores diferentes para retratar um mesmo cenário, o que não conferia credibilidade à informação gerada.

3.3 – Necessidades encontradas no processo de geração da informação

No entendimento do processo de geração da informação realizado pela empresa, foram mapeadas as seguintes necessidades:

- Automatização do processo dados;
- Confiabilidade nos dados gerados;
- Padronização das análises e *dashboards*;
- Disponibilização das bases de dados para download;
- Acesso via web dentro e fora da rede da empresa;
- Possibilidade de consumo por dispositivos móveis.

3.4 – Solução de BI proposta

A arquitetura da solução de BI proposta pode ser observada na figura 3.2. De forma geral, após o mapeamento das bases de dados utilizadas, o processo se inicia com a extração dos dados que estão no BD. Em seguida, é realizado o tratamento dos dados conforme as regras necessárias e, posteriormente, o carregamento na nuvem, onde são armazenados e utilizados para o desenvolvimento dos produtos de inteligência na plataforma de BI.

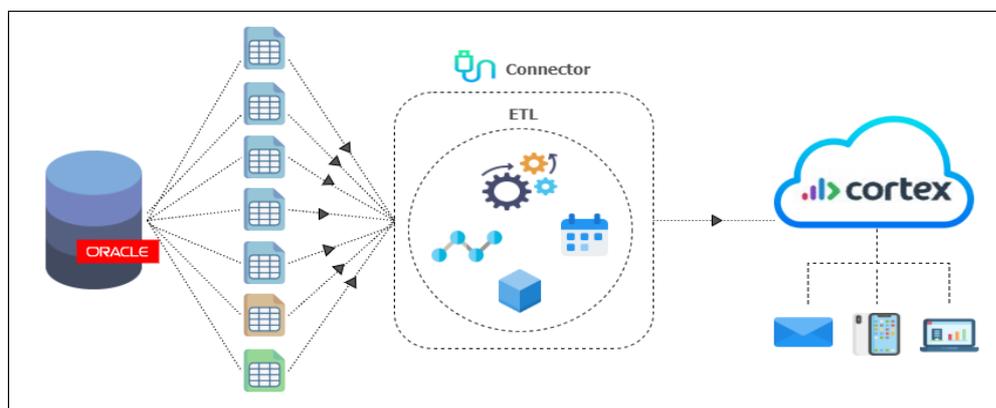


Figura 3.2 - Arquitetura de BI proposta.
Fonte: Elaborado pelo Autor.

3.4.1 – Fontes de dados

Para o desenvolvimento do projeto, foram mapeadas as fontes de dados necessárias, sendo ao todo sete bases: cinco referentes às fontes de receita da empresa, uma com a composição da carteira de clientes e uma referente a estrutura da área de vendas. As bases de dados utilizadas no processo são:

- Fonte de Receita 1 (FR1) e Fonte de Receita 2 (FR2) - apesar de serem bases distintas, FR1 e FR2 possuem a mesma estrutura e exibem dados da receita mensal por cliente por serviço contratado (Quadro 3.1).

Quadro 3.1 - Estrutura e descrição das bases de dados FR1 e FR2.

| Campo | Tipo | Exemplo | Descrição |
|-------------------------|--------|-------------|------------------------------------|
| DATA_MOV | TEXTO | 201711 | Data do arquivo |
| CONTA_CORRENTE | TEXTO | 000987 | Numero do cliente |
| SIGLA_SERVICO | TEXTO | BDL | Sigla do serviço |
| SERVICO | TEXTO | Banda Larga | Nome do serviço |
| SG_SISTEMA_FATM_SERVICO | TEXTO | ABC | Informação não utilizada em vendas |
| CD_MACRO | TEXTO | M12345 | Informação não utilizada em vendas |
| INTERLIGACAO | TEXTO | CDE345 | Informação não utilizada em vendas |
| NOME_INTERLIGACAO | TEXTO | D3456 | Informação não utilizada em vendas |
| DESIGNACAO | TEXTO | 12348765 | Informação não utilizada em vendas |
| FLAG_PGTO | TEXTO | P | Informação não utilizada em vendas |
| CD_PLANO | TEXTO | 929 | Informação não utilizada em vendas |
| VALOR | NÚMERO | 99,5 | Valor do serviço |

As linhas em verde correspondem aos campos selecionados na consulta para extração dos dados.

Fonte: Elaborado pelo Autor.

- Fonte de Receita 3 do ano atual (FR3_atual) e Fonte de Receita 3 do ano anterior (FR3_anterior) - ambas fazem referência a mesma fonte de receita e possuem mesma estrutura: exibem dados da receita mensal por cliente por serviço contratado, mas de anos diferentes. A FR3_atual possui dados do ano atual enquanto a FR3_anterior retrata dados do ano anterior (Quadro 3.2).

Quadro 3.2 - Estrutura e descrição das bases de dados FR3_atual e FR3_anterior.

| Campo | Tipo | Exemplo | Descrição |
|--------------------|--------|---------|---|
| CGC | TEXTO | 191 | Informação não utilizada em vendas |
| DIRNAC_FOTO_ANT | TEXTO | VES-4 | Informação não utilizada em vendas |
| DIRPARC_FOTO_ANT | TEXTO | VES-4 | Informação não utilizada em vendas |
| DIRNAC_FOTO_ATUAL | TEXTO | VES-4 | Informação não utilizada em vendas |
| DIRPARC_FOTO_ATUAL | TEXTO | VES-4 | Informação não utilizada em vendas |
| CONTA_CORRENTE | TEXTO | 123456 | Numero do cliente |
| TIPO_SERVICO | TEXTO | V | Tipo de serviço |
| SIGLA_SERVICO | TEXTO | V3456 | Sigla do serviço |
| SERVICO | TEXTO | Voz | Nome do serviço |
| CD_MACRO | TEXTO | C123 | Informação não utilizada em vendas |
| VALOR_JAN | NÚMERO | 99,89 | Valor do serviço em janeiro do ano da base gerada |
| VALOR_FEV | NÚMERO | 87,89 | Valor do serviço em fevereiro do ano da base gerada |
| VALOR_MAR | NÚMERO | 76,89 | Valor do serviço em março do ano da base gerada |
| VALOR_ABR | NÚMERO | 99,99 | Valor do serviço em abril do ano da base gerada |
| VALOR_MAI | NÚMERO | 102,25 | Valor do serviço em maio do ano da base gerada |
| VALOR_JUN | NÚMERO | 100,34 | Valor do serviço em junho do ano da base gerada |
| VALOR_JUL | NÚMERO | 80,54 | Valor do serviço em julho do ano da base gerada |
| VALOR_AGO | NÚMERO | 77,75 | Valor do serviço em agosto do ano da base gerada |
| VALOR_SET | NÚMERO | 89,99 | Valor do serviço em setembro do ano da base gerada |
| VALOR_OUT | NÚMERO | 90,91 | Valor do serviço em outubro do ano da base gerada |
| VALOR_NOV | NÚMERO | 80,72 | Valor do serviço em novembro do ano da base gerada |
| VALOR_DEZ | NÚMERO | 98,97 | Valor do serviço em dezembro do ano da base gerada |
| ANO | TEXTO | 2020 | Ano da receita gerada |

As linhas em verde correspondem aos campos selecionados na consulta para extração dos dados.

Fonte: Elaborado pelo Autor.

- Fonte de Receita 4 (FR4) - apresenta os dados de receita mensal por cliente por serviço contratado (Quadro 3.3).

Quadro 3.3 - Estrutura e descrição da base de dados FR4

| Campo | Tipo | Exemplo | Descrição |
|----------------|--------|----------------|------------------------------------|
| PERIODO | DATA | 20/17/01 | Data do arquivo |
| CONTA_CORRENTE | TEXTO | 123456 | Numero do cliente |
| CNPJ14 | TEXTO | 12387646578909 | CPJN cliente com 14 digitos |
| CNPJ8 | TEXTO | 34040183 | CPJN cliente com 8 digitos |
| CICLO | TEXTO | | Informação não utilizada em vendas |
| SIGLA_SERVICO | TEXTO | EDI | Sigla do serviço |
| SERVICO | TEXTO | Internet | Nome do serviço |
| PLANO | TEXTO | | Informação não utilizada em vendas |
| VALOR_BRUTO | NÚMERO | 749,84 | Valor bruto do serviço |
| VALOR_LIQUIDO | NÚMERO | 707,47 | Valor liquido do serviço |
| MAT_GC_NAC | TEXTO | 53627 | Informação não utilizada em vendas |
| NOME_GC_NAC | TEXTO | JOAO | Informação não utilizada em vendas |

As linhas em verde correspondem aos campos selecionados na consulta para extração dos dados.

Fonte: Elaborado pelo Autor.

- Base de Clientes da Empresa (Clientes) - exibe os dados dos clientes presentes na carteira da empresa, bem como os colaboradores responsáveis por cada cliente (Quadro 3.4). Vale destacar que a base de Clientes também possui os dados de *users*, para as posições de GEV e GC, utilizados nos sistemas da empresa. Essas informações serão importantes para a configuração de acessos aos dados na plataforma de BI.

Quadro 3.4 - Estrutura e descrição da base de dados Clientes

| Campo | Tipo | Exemplo | Descrição |
|-----------------------|-------|-----------------|--|
| GPR | TEXTO | A23 | Informação não utilizada em vendas |
| CNAE | TEXTO | 7586948 | Classificação da atividade do cliente |
| DV_TRATADA | TEXTO | D32 | Código da diretoria responsável pelo cliente |
| UF | TEXTO | RS | UF |
| CONTA15 | TEXTO | 273849586784938 | Numero do cliente |
| CONTA_PRINCIPAL_GRUPO | TEXTO | 273849586784938 | Número principal do grupo cliente |
| CONTA11 | TEXTO | 27384958678 | Número do grupo do cliente |
| NOME_RZ_SOC | TEXTO | EMPRESA A | Razão social do cliente |
| DE_TRATADA | TEXTO | DE1 | Diretoria executiva responsável pelo cliente |
| FAIXA_FATURAMENTO | TEXTO | S | Faixa de faturamento |
| CNPJ8 | TEXTO | 2315431 | CPJN cliente com 8 dígitos |
| CIDADE | TEXTO | PORTO ALEGRE | Cidade |
| NOME_DE | TEXTO | PEDRO | Diretor executivo responsável pelo cliente |
| CONTA_PRINCIPAL | TEXTO | S | Flag conta principal |
| NOME_GEV | TEXTO | JOAO | Gerente executivo responsável pelo cliente |
| CNPJ14 | TEXTO | 12387646578909 | CPJN cliente com 14 dígitos |
| VERTICAL | TEXTO | INDÚSTRIAS | Vertical do cliente |
| GRUPO_EMPRESARIAL | TEXTO | GRUPO A | Grupo do cliente |
| NOME_GC | TEXTO | JOANA | Gerente de contas responsável pelo cliente |
| NOME_DV | TEXTO | JOSE | Diretor de vendas responsável pelo cliente |
| CONTA_CONTROLADORA | TEXTO | Não | Flag conta controladora |
| ANO_MES_FECH | TEXTO | 201809 | Data de fechamento |
| USERID_GC | TEXTO | ARGONZO | user do gerente de contas responsável no sistema |
| USERID_GEV | TEXTO | VANLOP | user do gerente executivo responsável no sistema |

As linhas em verde correspondem aos campos selecionados na consulta para extração dos dados.

Fonte: Elaborado pelo Autor.

- Estrutura Comercial - o uso dessa base no processo é unicamente voltado para obter os *users* de acesso aos sistemas da empresa dos DEs, DVs e suportes comerciais (Quadro 3.5), uma vez que essas informações não estão na base de Clientes.

Quadro 3.5 - Estrutura e descrição da base de dados da Estrutura Comercial.

| Campo | Tipo | Exemplo | Descrição |
|--------------------|-------|------------|------------------------------------|
| DT_ARQUIVO | DATA | 10/10/2020 | Data do arquivo |
| MES_REF | TEXTO | 202010 | Mês de referência |
| DE | TEXTO | DE1 | Diretoria Executiva |
| DIR_EXEC | TEXTO | JOAO | Diretor Executivo |
| DC | TEXTO | C2334 | Informação não utilizada em vendas |
| DIR | TEXTO | T543 | Informação não utilizada em vendas |
| GRV | TEXTO | OT859 | Informação não utilizada em vendas |
| DIRETORIA | TEXTO | DV6 | Código da diretoria de Vendas |
| USERID_DV | TEXTO | PEDRO3 | Diretor de Vendas |
| GM | TEXTO | S | Informação não utilizada em vendas |
| NOME_AREA_RESUMIDO | TEXTO | AREA2 | Nome da área |
| USERID_DE | TEXTO | JOAO1 | User do Diretor de Vendas |
| USERID_SUORTE | TEXTO | JOSE2 | User do suporte comercial |

As linhas em verde correspondem aos campos selecionados na consulta para extração dos dados.

Fonte: Elaborado pelo Autor.

3.4.2 – Extração dos dados

As bases de dados mapeadas foram armazenadas em um banco de dados relacional com SGDB Oracle. Para o processo de extração, fez-se uso da linguagem de programação SQL. Seguindo as regras de negócio apresentadas pela área de vendas, os campos marcados nos quadros 3.1 a 3.5 foram selecionados no BD. Para as bases FR1, FR2 e FR4, foram filtrados os registros correspondentes ao período compreendido entre o ano anterior e o mês atual referente ao momento em que a extração foi realizada. As bases FR3_atual e FR3_anterior tiveram sua extração total, pois seus dados já compreendiam o período temporal desejado. Após a extração, os dados seguiram para a etapa de transformação.

3.4.3 – Transformação dos dados

Neste projeto, a transformação dos dados visa a consolidação das diferentes bases de receita em uma única base enriquecida com as informações de clientes e de acesso de usuários. A estrutura escolhida prioriza performance ao invés de ocupação do espaço em disco. Tal

decisão foi tomada com base em testes previamente realizados para o desempenho do carregamento de análises elaboradas dentro da plataforma Cortex, a ferramenta de BI utilizada na elaboração dos indicadores e dashboards.

O processo de transformação dos dados foi desenvolvido no *Pentaho Data Integration* (PDI). O PDI é um módulo da suíte de BI *open source* da Pentaho, que inclui ferramentas que auxiliam nas etapas do processo de decisão de negócios. Seu grande diferencial está no conjunto poderoso de utilidades de ETL que oferece, possibilitando ao usuário criar projetos complexos com uma interface gráfica amigável (Figura 3.3), sem a necessidade de uma linha de código (ALOOMA TEAM, 2018; SPEC INDIA, 2021).

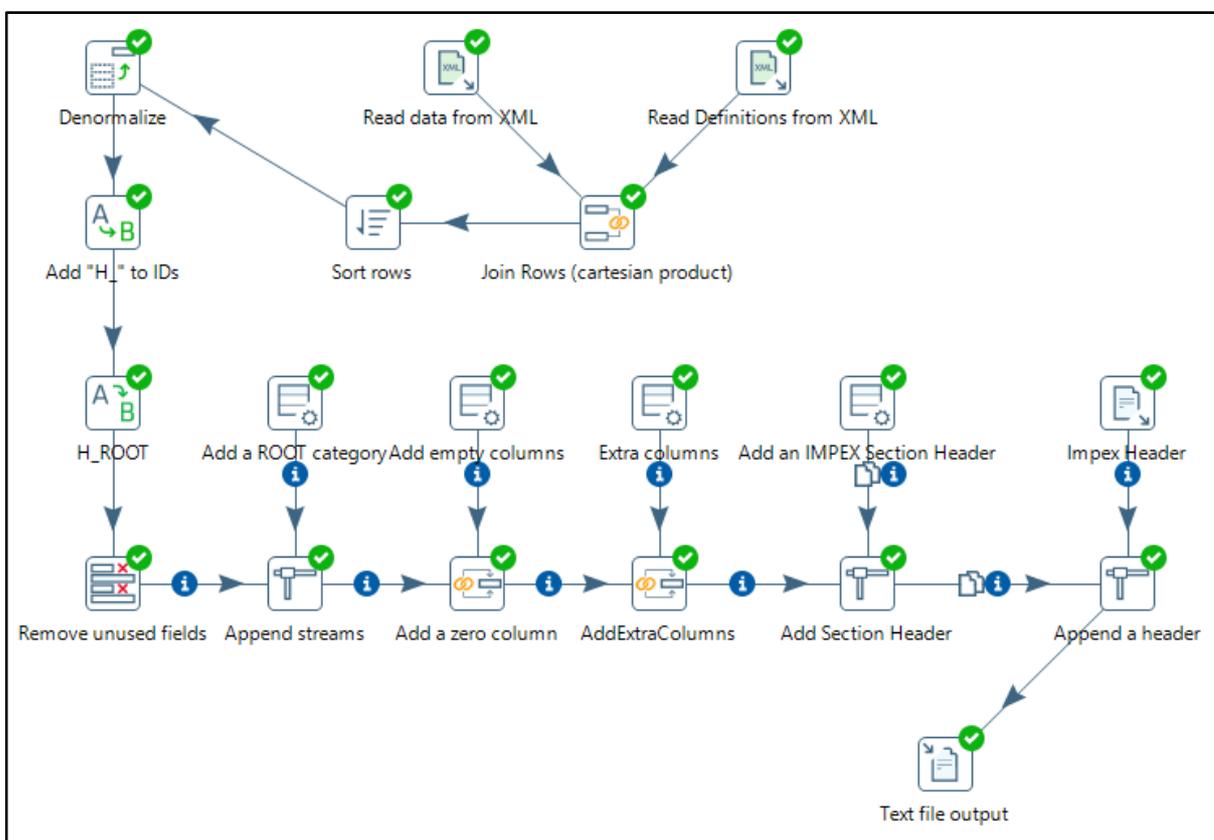


Figura 3.3 - Exemplo de um fluxo de ETL desenvolvido no PDI
 Fonte: ALIEV, 2017

Na primeira parte do tratamento de dados, as bases de receita são extraídas do BD e passam por algumas transformações (Figura 3.4) com o objetivo de padronizar sua estrutura, possibilitando a concatenação de todas em uma base total (FR_Total).

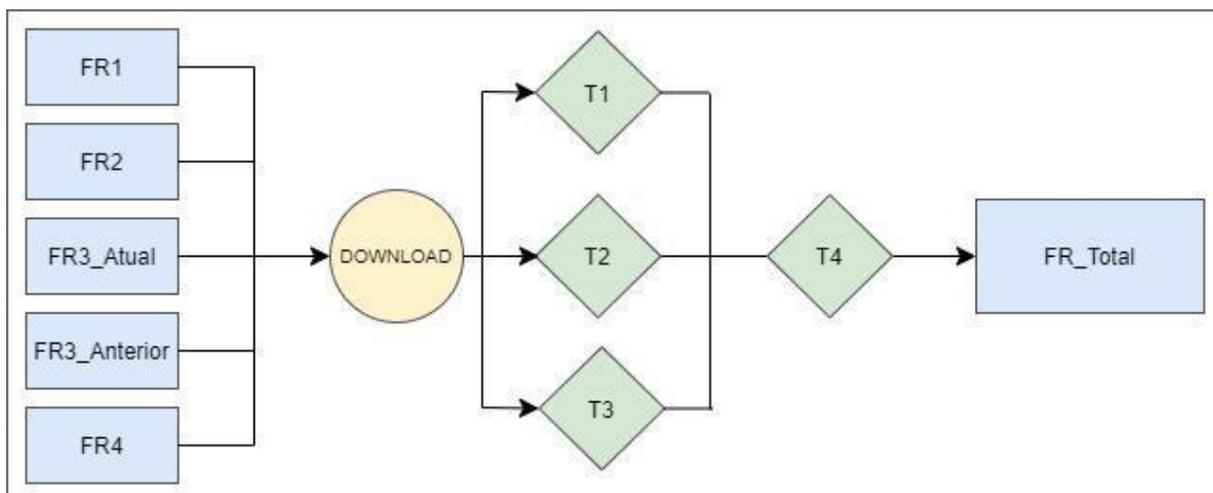


Figura 3.4 - Fluxo de transformação 1: ajuste e concatenação das bases de receita.

Fonte: Elaborado pelo Autor.

- T1:
 1. Criar campo 'FONTE' na base FR1 e preencher com 'FR1';
 2. Criar campo 'FONTE' na base FR2 e preencher com 'FR2';
 3. Concatenar as bases FR1 e FR2;
 4. Converter o campo 'DATA_MOV' para tipo 'data' como o padrão '01/mm/aaa'.
 5. Renomear o campo 'DATA_MOV' para 'PERIODO'.
- T2:
 1. Concatenar as bases FR3_Atual e FR3_Anterior;
 2. Criar campo 'FONTE' e preencher com 'FR3';
 3. Criar campo tipo 'data' com o nome 'PERIODO';
 4. Criar campo tipo 'número' com nome 'VALOR';
 5. Reestruturar os 12 campos de valor de receita referentes a cada mês em um único campo, transformando as colunas de valor em linhas.
 6. Durante a reestruturação, preencher o campo 'VALOR', criado anteriormente, com os valores oriundos das colunas de valor;
 7. Durante a reestruturação, preencher o campo 'PERIODO' criado anteriormente, com os respectivos meses indicados nos campos de valor e com a informação do campo 'ANO', seguindo o padrão '01/mm/aaaa'.
- T3:
 1. Criar campo 'FONTE' na base FR4 e preencher com 'FR4';
 2. Renomear o campo 'VALOR_BRUTO' da base FR4 para 'VALOR';

- T4:
 1. Concatenar a resultante das bases FR1, FR2, FR3_Atual, FR3_Anterior e FR4 em uma única base de receita que possui a indicação de suas fontes no campo ‘FONTE’ criado nos steps anteriores e salvar como FR_Total.

A segunda etapa compreende o processo de enriquecimento da base de clientes com a informação de *users* (usuários de acesso aos sistemas da empresa) dos DVs e DEs (Figura 3.5).

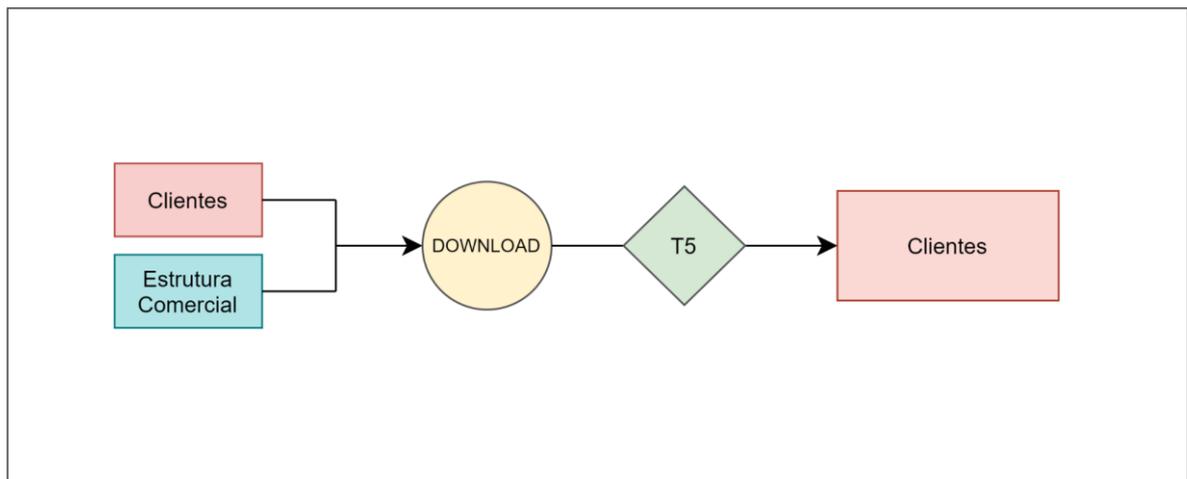


Figura 3.5 - Fluxo de transformação 2: enriquecimento da base de clientes.
Fonte: Elaborado pelo Autor.

- T5:
 1. Realizar o *left join* entre as bases Clientes e Estrutura Comercial pelos campos ‘DV_TRATADA’ e ‘DIRETORIA’;
 2. Retornar os campos ‘USERID_DV’, ‘USERID_DE’ E ‘USERID_SUPORTE’ e enriquecer a base Clientes.
 3. Salvar a nova base Clientes.

A última etapa de transformação consiste no enriquecimento da FR_Total com os dados de clientes e usuários (Figura 3.6). Durante o desenvolvimento do processo, o ponto focal da área de vendas da empresa solicitou um complemento na estrutura dos dados: a inserção de dois campos, ‘FAMILIA’ e ‘PLATAFORMA’ correspondentes a agregações dos serviços que eram vendidos pela empresa. Para atender a solicitação, foi disponibilizada uma base (de/para Serviços) com as correlações entre o campo ‘SIGLA_SERVIÇO’ existente na base FR_Total e os campos ‘FAMILIA’ e ‘PLATAFORMA’ os quais deveriam ser criados (Quadro 3.6).

Quadro 3.6 - Estrutura da base de serviços por famílias e plataformas.

| SIGLA_SERVICO | FAMILIA | PLATAFORMA |
|---------------|-----------|--------------|
| A01 | FAMILIA A | PLATAFORMA 1 |
| A02 | FAMILIA A | PLATAFORMA 1 |
| A03 | FAMILIA A | PLATAFORMA 1 |
| B01 | FAMILIA B | PLATAFORMA 1 |
| B02 | FAMILIA B | PLATAFORMA 1 |
| B03 | FAMILIA B | PLATAFORMA 1 |
| C01 | FAMILIA C | PLATAFORMA 2 |
| C02 | FAMILIA C | PLATAFORMA 2 |
| D01 | FAMILIA D | PLATAFORMA 2 |
| E01 | FAMILIA E | PLATAFORMA 3 |
| E02 | FAMILIA E | PLATAFORMA 3 |
| E03 | FAMILIA E | PLATAFORMA 3 |

Fonte: Elaborado pelo Autor.

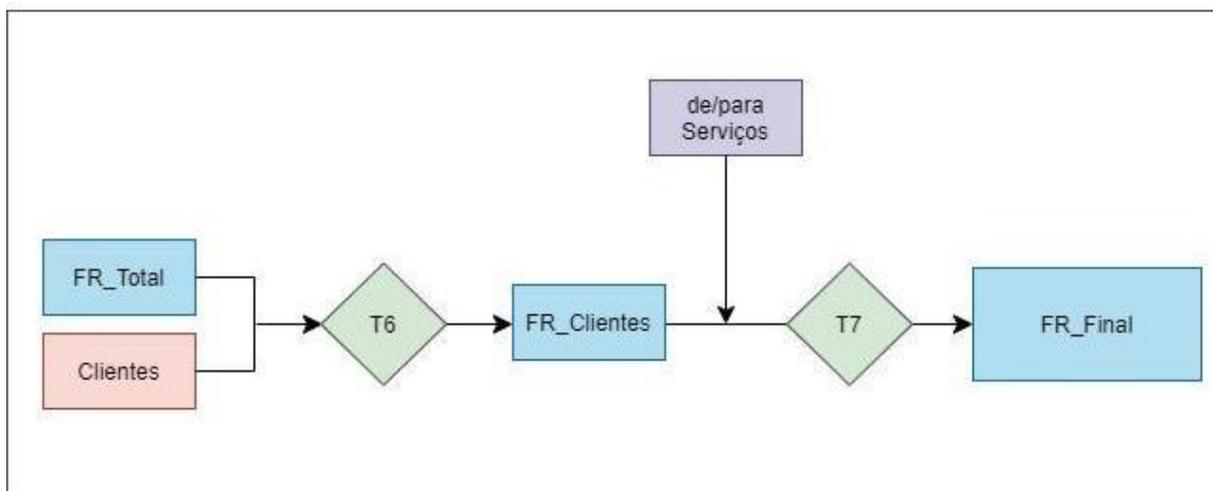


Figura 3.6 - Fluxo de transformação 3: enriquecimento da base FR_Total.

Fonte: Elaborado pelo Autor.

- T6:
 4. Realizar o *left join* entre as bases FR_Total e Clientes pelos campos ‘CONTA_CORRENTE’ e ‘CONTA15’;
 5. Retornar todos os campos da base de Clientes e enriquecer a base FR_Total.
- T7:
 6. Realizar o *left join* entre as bases FR_Clientes e de/para Serviços pelo campo ‘SIGLA_SERVIÇO’;
 7. Criar um campo multivalorado chamado ‘CONTROLE_ACESSO’ concatenando os campos ‘USERID_DE’, ‘USERID_DV’, ‘USERID_GEV’,

‘USERID_GC’ e ‘USERID_SUPORTE’ da base FR_Clientes, separando-os por ‘|’. Exemplo: user1|user2|user3|user4|user5

8. Enriquecer a base FR_Clientes com as informações de ‘PLATAFORMA’ e ‘FAMILIA’ e salvar como FR_Final (Tabela 3.7).

Quadro 3.7 - Base FR_Final: estrutura da versão final gerada no tratamento dos dados.

| Campo | Tipo | Exemplo | Descrição |
|-----------------------|------------|-----------------------------------|--|
| DATA_MOV | DATA | 01/10/2020 | Data do arquivo |
| CONTA_CORRENTE | TEXTO | 000987 | Numero do cliente |
| SIGLA_SERVICO | TEXTO | BDL | Sigla do serviço |
| SERVICO | TEXTO | Banda Larga | Nome do serviço |
| VALOR | NÚMERO | 99,5 | Valor do serviço |
| DV_TRATADA | TEXTO | D32 | Código da diretoria responsável pelo cliente |
| UF | TEXTO | RS | UF |
| CONTA15 | TEXTO | 273849586784938 | Numero do cliente |
| CONTA_PRINCIPAL_GRUPO | TEXTO | 273849586784938 | Número principal do grupo cliente |
| CONTA11 | TEXTO | 27384958678 | Número do grupo do cliente |
| NOME_RZ_SOC | TEXTO | EMPRESA A | Razão social do cliente |
| DE_TRATADA | TEXTO | DE1 | Diretoria executiva responsável pelo cliente |
| CIDADE | TEXTO | PORTO ALEGRE | Cidade |
| NOME_DE | TEXTO | PEDRO | Diretor executivo responsável pelo cliente |
| CONTA_PRINCIPAL | TEXTO | S | Flag conta principal |
| NOME_GEV | TEXTO | JOAO | Gerente executivo responsável pelo cliente |
| CNPJ14 | TEXTO | 12387646578909 | CPJN cliente com 14 dígitos |
| VERTICAL | TEXTO | INDÚSTRIAS | Vertical do cliente |
| GRUPO_EMPRESARIAL | TEXTO | GRUPO A | Grupo do cliente |
| NOME_GC | TEXTO | JOANA | Gerente de contas responsável pelo cliente |
| NOME_DV | TEXTO | JOSE | Diretor de vendas responsável pelo cliente |
| CONTA_CONTROLADORA | TEXTO | Não | Flag conta controladora |
| ANO_MES_FECH | TEXTO | 201809 | Data de fechamento |
| USERID_GC | TEXTO | ARGONZO | user do gerente de contas responsável no sistema |
| USERID_GEV | TEXTO | VANLOP | user do gerente executivo responsável no sistema |
| USERID_DV | TEXTO | PEDRO3 | Diretor de Vendas |
| USERID_DE | TEXTO | JOAO1 | User do Diretor de Vendas |
| USERID_SUPORTE | TEXTO | JOSE2 | User do suporte comercial |
| CONTROLE_ACESSO | MULTIVALOR | ARGONZO VANLOP PEDRO3 JOAO1 JOSE2 | Campo que concatena todos os users separados por ' ' |

Fonte: Elaborado pelo Autor.

3.4.4 – Carga dos dados

O processo de carga dos dados acontece em seguida ao de transformação. O projeto de ETL desenvolvido no PDI gera uma pasta contendo os arquivos referentes às transformações do processo. Esses arquivos são compactados e inseridos no Connector, ferramenta que faz a conexão entre a fonte e o destino dos dados, para a execução do processo de ETL. Neste trabalho, foi utilizado o Cortex Connector, uma ferramenta para carga de dados que opera em conjunto com a plataforma de BI Cortex. A interface do Connector é bem simples e consiste basicamente em duas sessões: *Data Sources* e *Data Connections* (Figura 3.7).

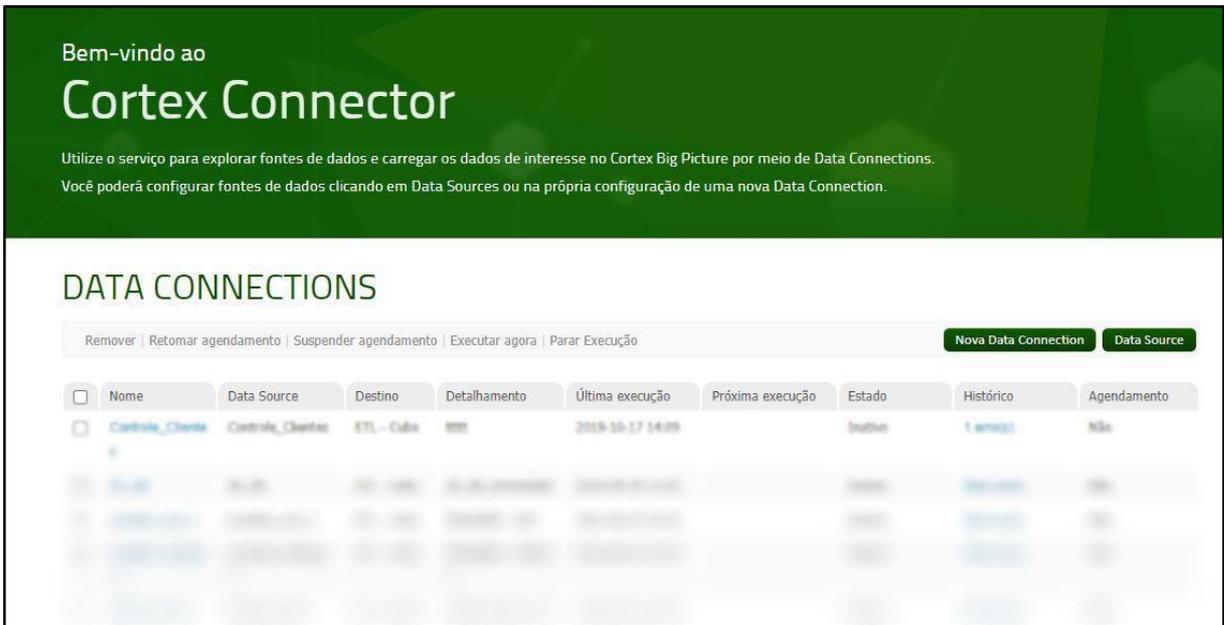


Figura 3.7 - Tela inicial do Cortex Connector.
 Fonte: Elaborado pelo autor no software Cortex Connector.

Na sessão *Data Sources* são configuradas as fontes onde o Connetor irá buscar os dados a serem tratados e carregados. Podem existir várias *data sources* configuradas no Connector e cada *data source* pode conter diversas *data connections*. No caso de uma *data source* do tipo projeto *kettle* (utilizado no desenvolvimento do presente trabalho), toda a especificação das fontes de dados, transformações e output está no arquivo que é gerado pelo PDI, bastando apenas carregar o mesmo no Connector (Figura 3.8).

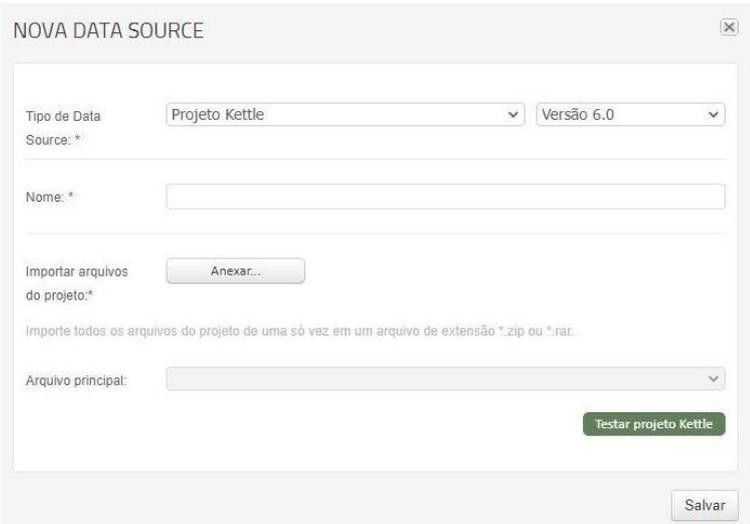


Figura 3.8 - Tela de configuração de uma data source tipo Projeto Kettle
 Fonte: Elaborado pelo autor no software Cortex Connector.

Após a configuração da *data source* cria-se a *data connection*, que é responsável pela conexão entre a *data source* e a plataforma de BI. Os passos para configurar uma *data connection* são:

1. Selecionar uma *data source* criada anteriormente (Figura 3.9);



Figura 3.9 - Configuração de uma *data connection*: selecionando uma *data source*.
Fonte: Elaborado pelo autor no software Cortex Connector.

2. Indicar o *Data Set* que receberá os dados dentro da plataforma de BI (Figura 3.10);



Figura 3.10 - Configuração de uma *data connection*: selecionando o *dataset*.
Fonte: Elaborado pelo autor no software Cortex Connector.

3. Agendar a recorrência da execução do processo (Figura 3.11).

NOVA DATA CONNECTION

1 Seleccione a fonte de dados 2 Autentique-se na plataforma 3 Configure a carga de dados 4 Agende a execução

AGENDE A EXECUÇÃO DA DATA CONNECTION

Nome Data Connection: *

Sem agendamento

A cada dia às

Domingo Segunda Terça Quarta Quinta Sexta Sábado

Cron expression

Utilizar todos os dados disponíveis. Utilizar dados a partir de: yyyy-mm-dd hh:mm:ss

Suspender o agendamento em caso de falha. Tempo máximo de extração segundos

Voltar Salvar

Figura 3.11 - Configuração de uma *data connection*: agendando a execução.
Fonte: Elaborado pelo autor no software Cortex Connector.

Após a criação da *data source*, da configuração e execução da *data connection*, os dados foram carregados na plataforma de BI Cortex.

3.4.5 – Armazenamento dos dados

Os dados carregados na plataforma Cortex ficam armazenados em data sets (cubos) que funcionam como as fontes de dados para as análises que serão desenvolvidas dentro da própria ferramenta de BI. A plataforma Cortex é uma ferramenta 100% em nuvem, assim como os dados que ficam nos data sets. Tais dados são armazenados em DW em nuvem com banco de dados OLAP *Snowflake*, que é uma tecnologia de BD otimizada para consulta e relatórios que usa estruturas multidimensionais para fornecer acesso rápido aos dados para análise. Os dados OLAP são organizados hierarquicamente e armazenados em cubos, em vez de tabelas. Esse tipo de estrutura facilita consultas de BI (MICROSOFT, [s.d.]).

3.4.6 – Configuração de acesso aos registros

Visando cumprir as regras de visualização dos dados conforme a hierarquia da área de vendas apresentada na página 15, foi elaborado um racional de segurança dos registros da base FR_Final (Tabela 3.7) sobre o campo CONTROLE_ACESSO (criado na etapa T7 de

transformação dos dados). O campo em questão é multivalorado e retorna uma combinação dos usuários de acesso aos sistemas da empresa. Esses usuários de acesso são de todos os responsáveis pelo cliente exibido em cada registro da base FR_Final, dentro da hierarquia de vendas (Tabela 3.8). Vale ressaltar que os logins dos usuários na plataforma de BI foram criados conforme os *users* contidos na tabela FR_Final (Tabela 3.7), para que fosse possível implementar o racional de segurança desejado.

Quadro 3.8 - Exemplo da composição do campo ‘CONTROLE_ACESSO’.

| DATA_MOV | SERVICO | VALOR | NOME_RZ_SOC | USERID_GC | USERID_GEV | USERID_DV | USERID_DE | USERID_SUPORTE | CONTROLE_ACESSO |
|------------|---------|--------|-------------|-----------|------------|-----------|-----------|----------------|--------------------------------|
| 01/10/2020 | BL | 120,99 | Cliente 1 | idGC1 | idGEV1 | idDV1 | idDE1 | idSP1 | idGC1 idGEV1 idDV1 idDE1 idSP1 |
| 01/10/2020 | VOZ | 29,90 | Cliente 2 | idGC2 | idGEV1 | idDV1 | idDE1 | idSP1 | idGC2 idGEV1 idDV1 idDE1 idSP1 |
| 01/10/2020 | IOT | 240,25 | Cliente 3 | idGC3 | idGEV2 | idDV2 | idDE1 | idSP1 | idGC3 idGEV2 idDV2 idDE1 idSP1 |
| 01/10/2020 | BL | 109,90 | Cliente 4 | idGC4 | idGEV3 | idDV3 | idDE2 | idSP2 | idGC4 idGEV3 idDV3 idDE2 idSP2 |

Fonte: Elaborado pelo Autor.

Dentro da plataforma Cortex existe uma função chamada *row level security* (RLS) que funciona restringindo o acesso aos registros de um *dataset* (base de dados carregada na plataforma) com base em um racional determinado. Fazendo uso do RLS foi desenvolvida a seguinte regra: se o *login* do usuário que está acessando a plataforma estiver contido no campo ‘CONTROLE_ACESSO’, o mesmo pode visualizar o registro correspondente, caso contrário, o registro não poderá ser visualizado. Na prática, essa regra retorna quantidades diferentes de registros de uma mesma base para usuários diferentes, garantindo que cada usuário visualize apenas os registros que compõem a receita de seus clientes.

3.4.7 – Desenvolvimento dos dashboards

Para o desenvolvimento das análises e dashboards, foi utilizada a plataforma Cortex, uma ferramenta de BI e Big Data, que dentre diversas outras funcionalidades, possui um conjunto de recursos para análise de dados (Figura 3.12), construção de indicadores e de *dashboards* interativos e customizáveis (Figura 3.13).

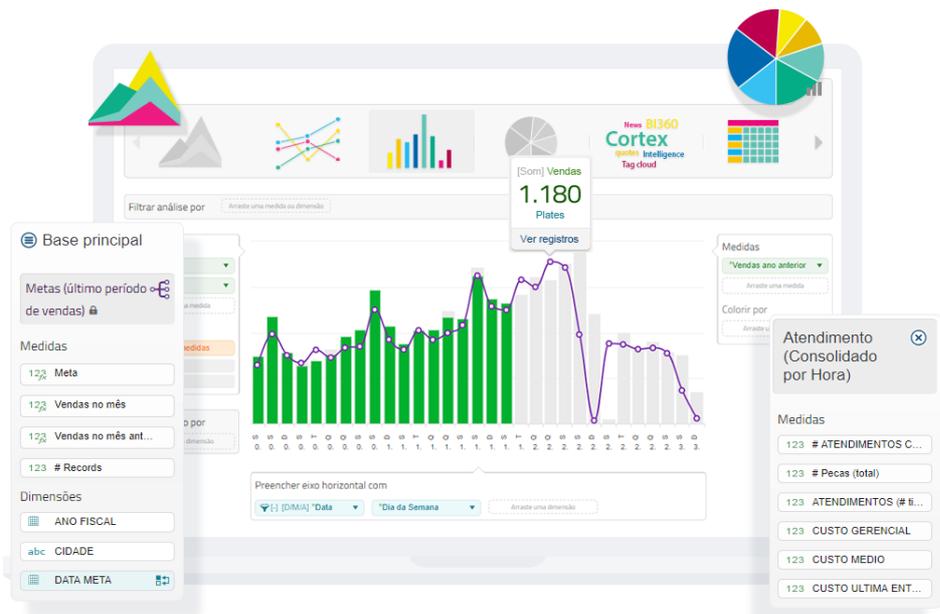


Figura 3.12 - Plataforma Cortex: interface para elaboração de análises.
Fonte: CORTEX INTELLIGENCE, 2021.

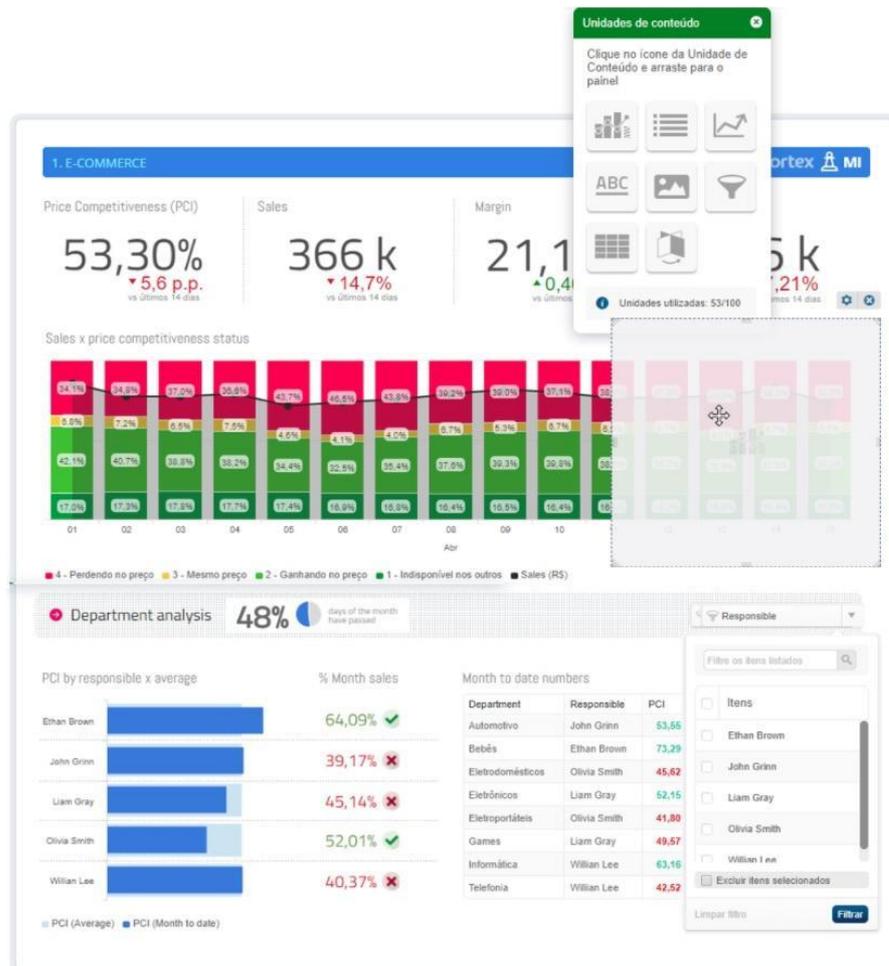


Figura 3.13 - Plataforma Cortex: interface para elaboração de dashboards.
Fonte: CORTEX INTELLIGENCE, 2021.

Entendidas as necessidades de informação da área, foram elaborados *dashboards* para consumo via notebooks, *desktops* e tablets, além de telas para consumo *mobile* via smartphones. Durante o desenvolvimento, as seguintes premissas foram consideradas:

- Exibir o resultado consolidado de vendas possibilitando o detalhamento (*drill down*) pela hierarquia da área de vendas: Presidente > DEs > DVs > GEVs > GCs;
- Exibir o resultado consolidado de vendas possibilitando o detalhamento (*drill down*) pela hierarquia dos serviços oferecidos pela empresa: plataforma de serviços > famílias de serviços > serviços;
- Exibir o resultado de vendas por clientes, possibilitando seleção de clientes específicos e indicando os maiores/menores geradores de receita;
- Exibir o resultado de vendas por verticais mostrando sua composição e possibilitando a listagem e detalhamentos dos clientes que as compõem;
- Possibilitar alternância entre a visão de mês atual e de ano contra ano (YTD) para todas as visões listadas.

Para atender essas necessidades, foram utilizados vários tipos de visualização de dados, tais como: tabelas, métricas, imagens condicionais, gráficos de barras simples, gráficos de barras 100%, gráficos setoriais (pizza), gráficos de linhas, gráficos combinados de linhas e barras, e gráficos de dispersão.

Capítulo 4

Resultados Obtidos

4.1 – Dashboards gerados no processo

Todo o processo descrito neste trabalho foi desenvolvido com o objetivo de viabilizar a construção de uma plataforma para a análise dos resultados de vendas da empresa apresentada. O conjunto de dashboards, desenvolvidos com base nas premissas expostas na página 33 apresenta visões padronizadas que se comunicam, de modo que o usuário possa navegar pelas telas herdando informações de filtros anteriores e detalhando-as em níveis menores de informação.

Todas as telas desenvolvidas exibem informações personalizadas conforme a posição do usuário dentro da hierarquia de vendas, cadastrada na criação do acesso do mesmo a plataforma. Os valores exibidos também são diferentes para cada usuário, visto que a configuração de acesso dos dados (sessão 3.4.6) restringe a visualização os registros da base de dados na plataforma conforme o perfil do usuário, exibindo somente seus resultados individuais e os de seus diretos.

4.1.1 – Receita por diretorias

O *dashboard* de Receita por Diretorias possibilita ao usuário entender a composição da receita da empresa através da hierarquia da área de vendas, que pode ser alterada entre as visões de mês atual e ano atual. A tela apresentada na figura 4.1 reflete a visão do presidente, exibindo o total da receita da empresa com suas variações e o seu detalhamento pelas diretorias executivas. A funcionalidade do *drill down* possibilita que o valor da receita seja detalhado em níveis menores de informação. Dessa forma, o resultado de uma DE pode ser detalhado pelo resultado de seus DVs, que pode ser detalhado por seus GEVs e GCs.

Além de exibir as informações de receita pelos níveis da hierarquia de vendas dentro do próprio dashboard, o usuário pode utilizar o resultado de seus filtros e *drills* para analisar o conjunto de informações desejado em um outro *dashboard*, por exemplo, entender o motivo da receita do diretor 21 ter caído 10% em relação ao último mês, detalhando essa informação no *dashboard* de receita por serviços ou então por clientes.

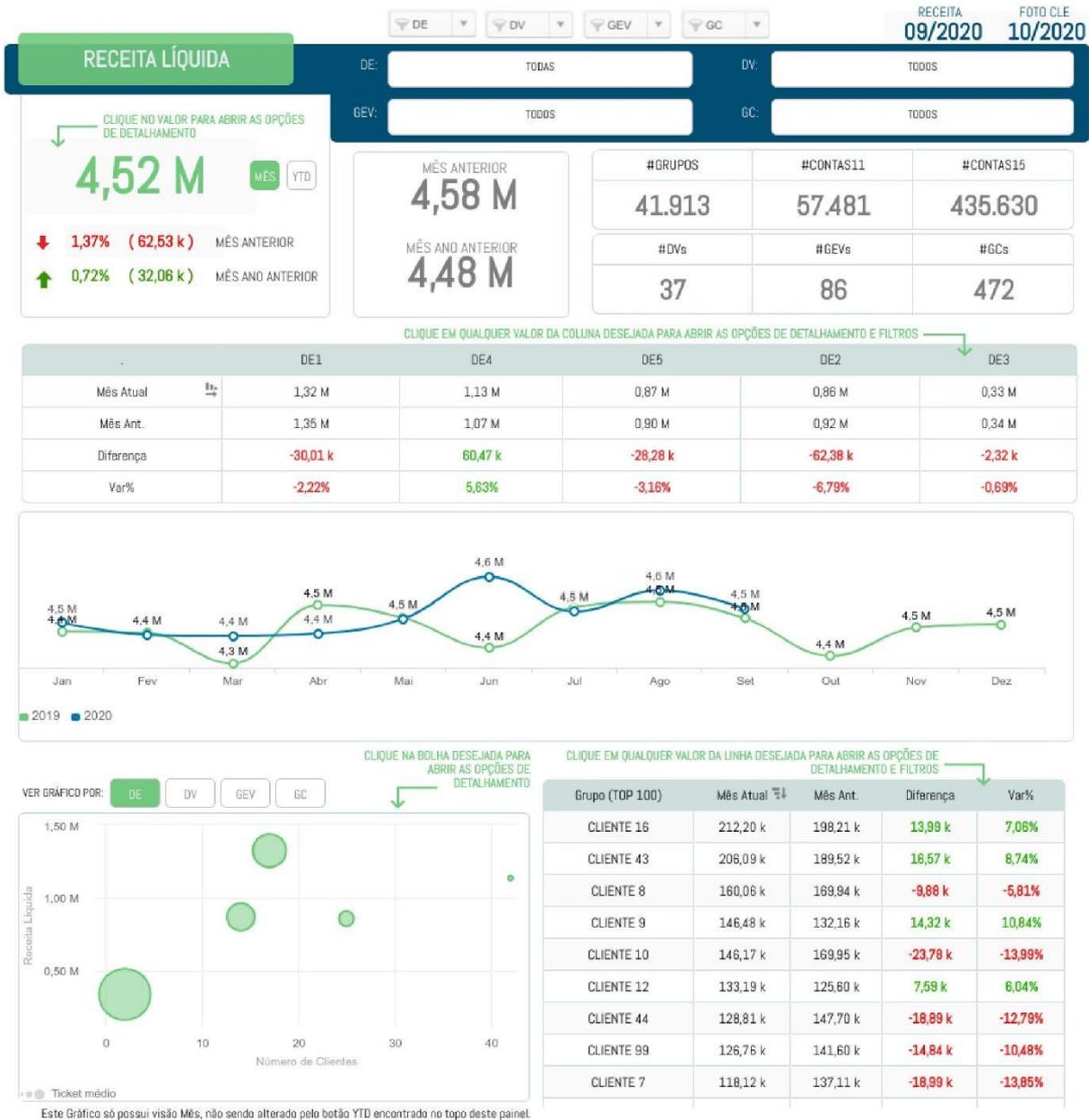


Figura 4.1 – Dashboard desenvolvido: Receita por diretorias.

Fonte: Elaborado pelo Autor.

4.1.2 – Receita por plataformas

Informações de receita por plataformas podem ser acessadas na tela apresentada na figura 4.2. Seguindo a identidade visual definida em todos os *dashboards* desenvolvidos, o usuário pode utilizar filtros para definir um escopo específico de dados a serem analisados ou então partir de um valor total e detalhar o mesmo através dos *drills*, passando pelas visões de plataformas, sua abertura por famílias até chegar à composição da receita por cada serviço oferecido pela empresa.

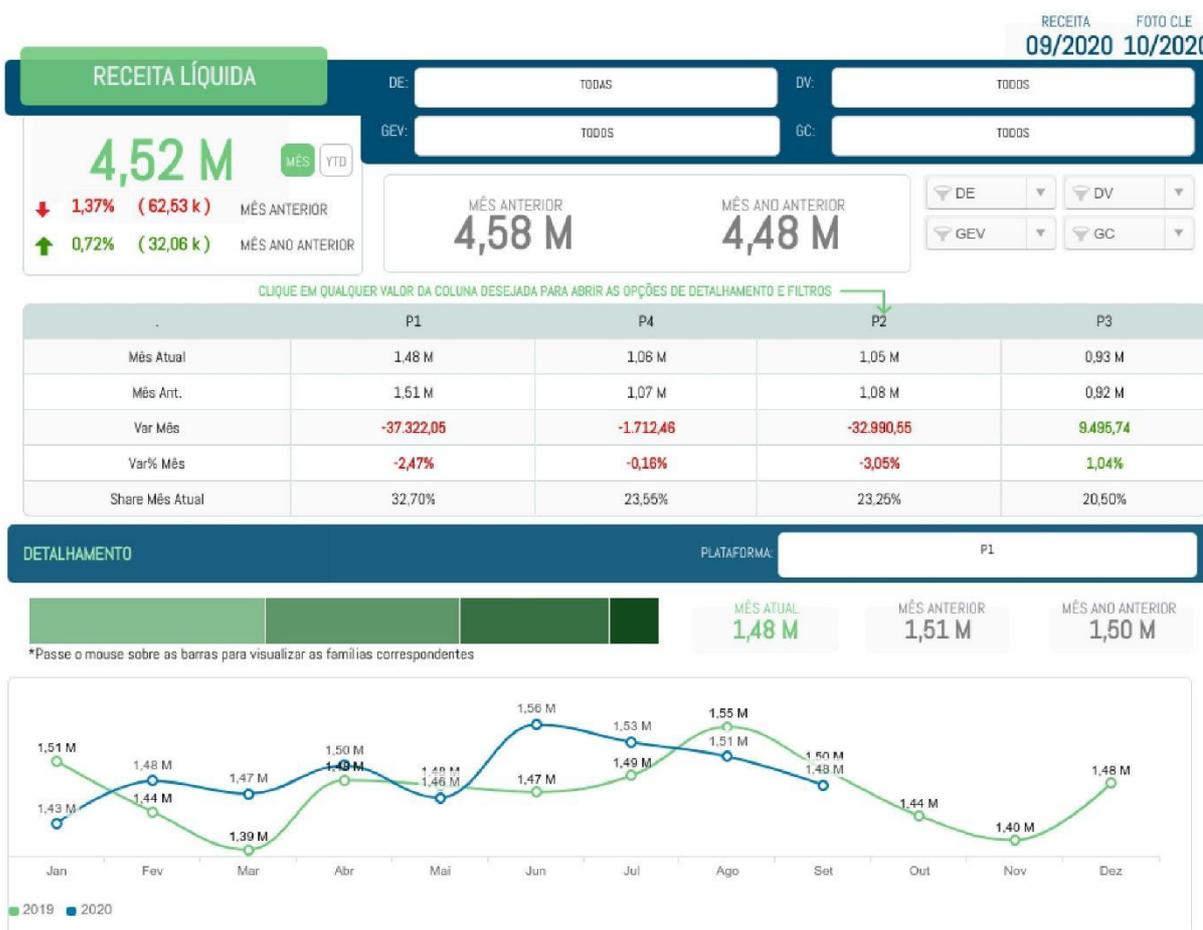


Figura 4.2 – Dashboard desenvolvido: Receita por plataformas.
Fonte: Elaborado pelo Autor.

4.1.3 – Receita por famílias

O dashboard de Receita por Famílias (figura 4.3) exibe o dado da receita agregado por famílias de serviços e apresenta uma estrutura visual semelhante ao de plataformas (figura 4.2), possibilitando o usuário a realizar filtros e detalhar a informação por serviços.

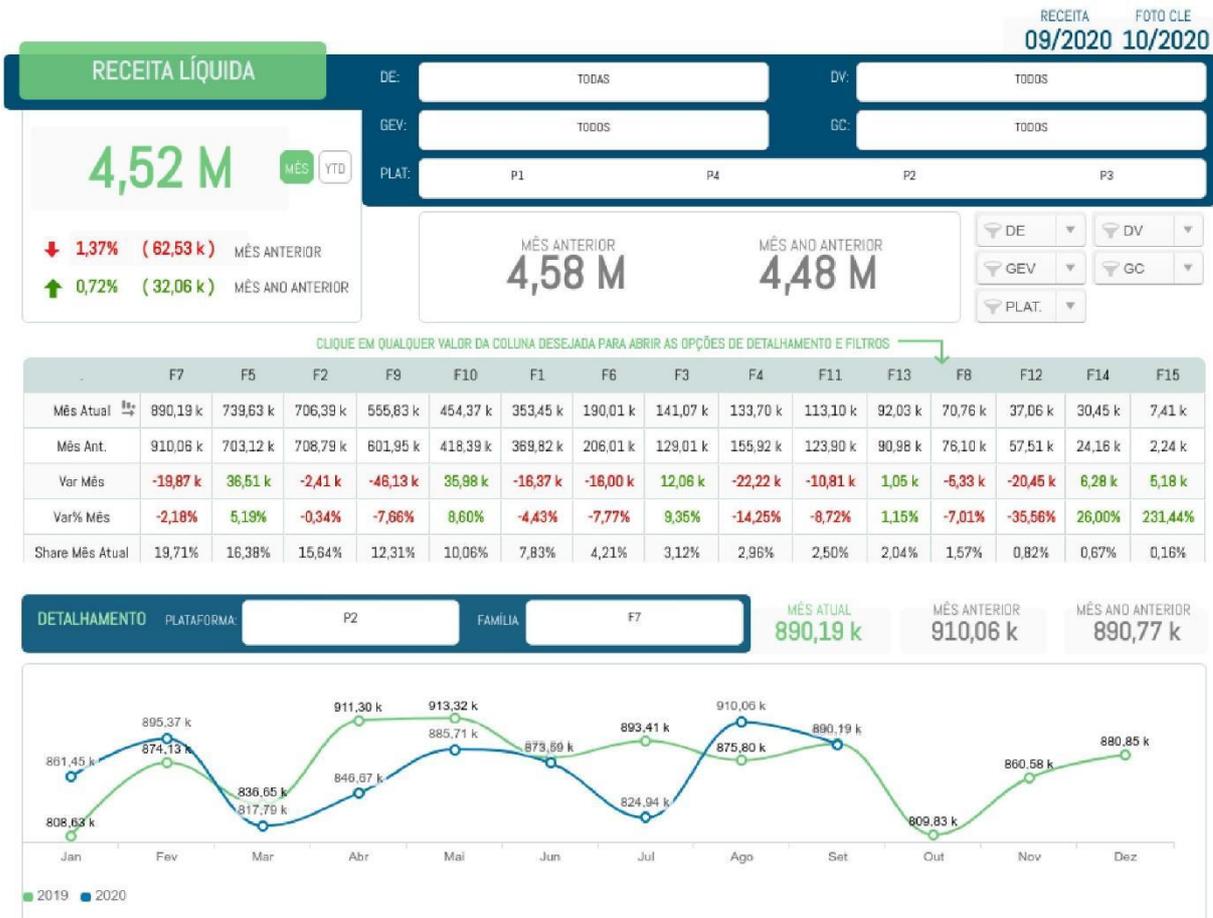


Figura 4.3 – Dashboard desenvolvido: Receita por famílias.
 Fonte: Elaborado pelo Autor.

4.1.4 – Receita por serviços

No dashboard de Receita por Serviços (figura 4.4) apresenta-se a distribuição da receita pelos serviços oferecidos pela empresa. Como o nível de serviços é o de menor granularidade na base de dados, não existe a possibilidade de se utilizar *drill down* nesse dashboard, porém não impossibilita o usuário de detalhar a receita de um determinado serviço por clientes ou verticais, por exemplo, utilizando a função de *drill* em outros *dashboards*.



Figura 4.4 – Dashboard desenvolvido: Receita por serviços.

Fonte: Elaborado pelo Autor.

4.1.5 – Receita por verticais

O dashboard de Receita por Verticais (figura 4.5) tem como objetivo detalhar a distribuição e evolução da receita conforme os setores de atuação dos clientes da empresa. Assim como as demais telas desenvolvidas, possui filtros e *drills* na primeira sessão, que facilitam o consumo e entendimento das informações por parte dos usuários. Na segunda sessão retorna o detalhamento da vertical selecionada pelo usuário, e o ranking dos dez maiores e menores clientes geradores de receita dentro da vertical em questão.

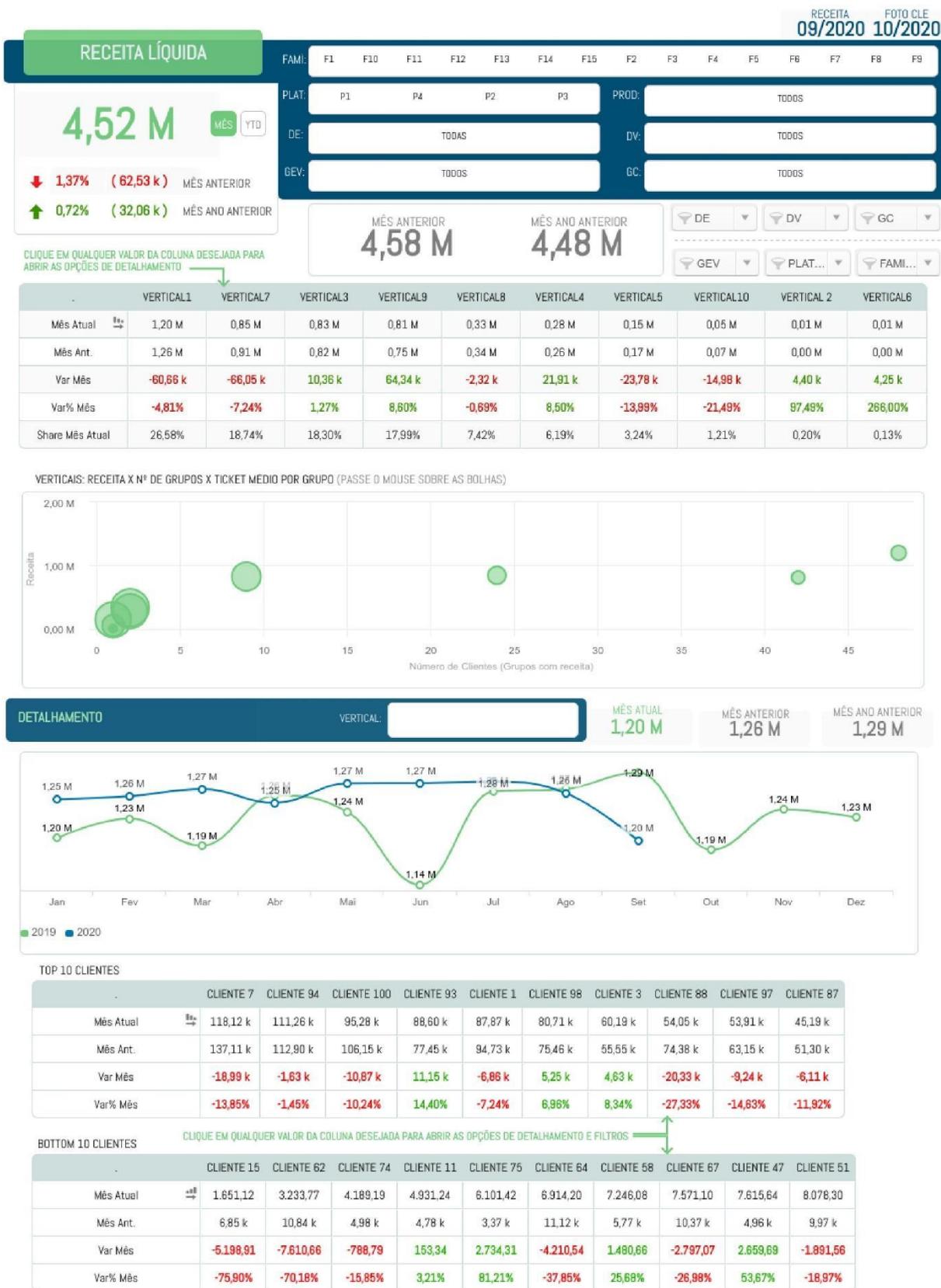


Figura 4.5 – Dashboard desenvolvido: Receita por verticais.
Fonte: Elaborado pelo Autor.

4.1.6 – Receita por clientes

No dashboard de Receita por Clientes (figura 4.6) o usuário é capaz de entender a distribuição da receita dentro de sua carteira de clientes. Por meio de *drills* para outros *dashboards*, pode visualizar os serviços consumidos por seus clientes, a evolução de suas receitas ao longo dos últimos anos e sua posição dentro da vertical na qual está inserido.

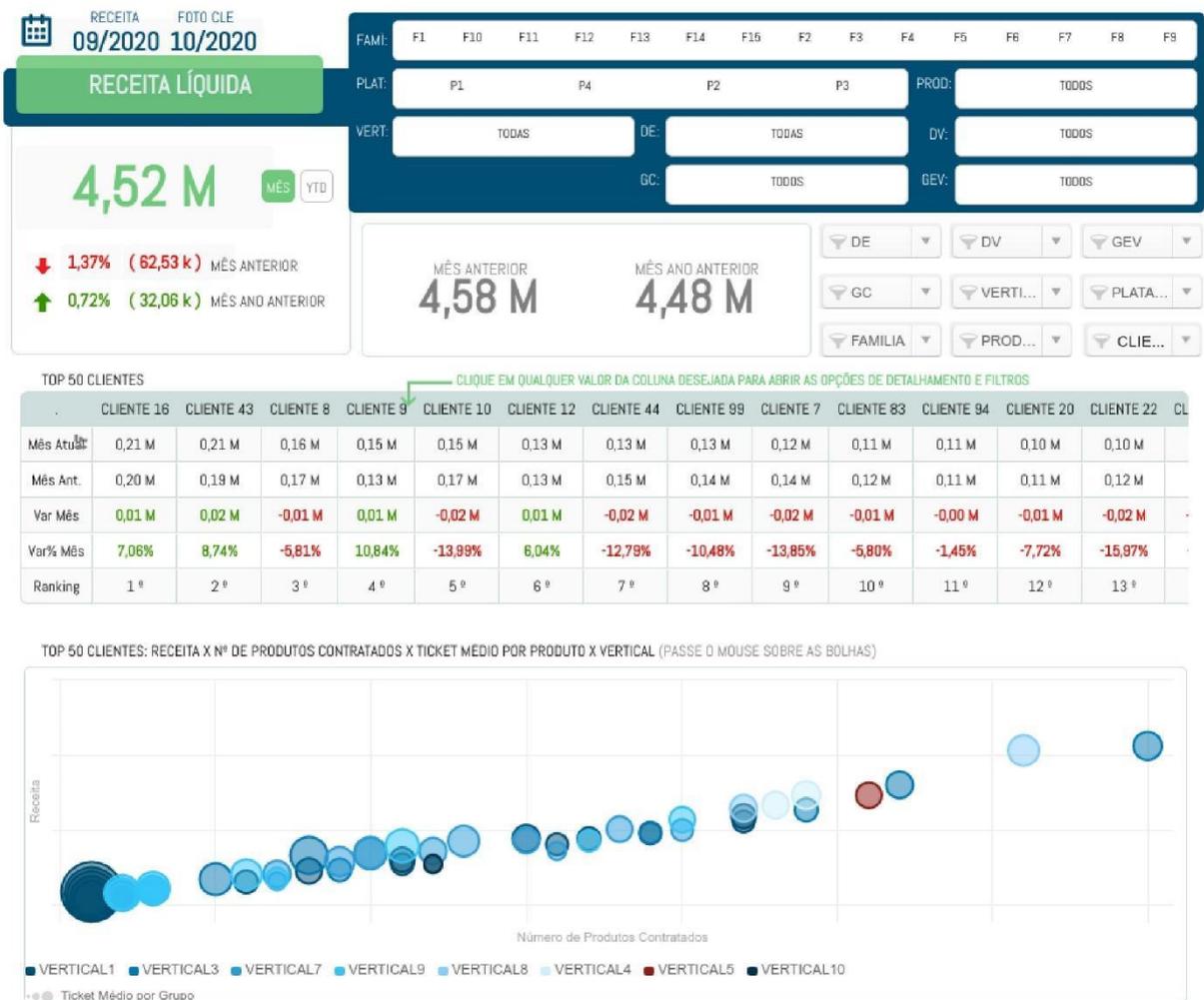


Figura 4.6 – Dashboard desenvolvido: Receita por clientes.
Fonte: Elaborado pelo Autor.

4.1.7 – Dashboards Mobile

Uma das solicitações da área de vendas era a necessidade de consumir os *dashboards* em dispositivos móveis, principalmente para suprir necessidade de informações para as equipes de vendas externas e consultar dados de forma rápida e prática. Para atender esta demanda, foram criados *dashboards* (figura 4.7) para consumo em smartphones, com as mesmas visões apresentadas nas seções anteriores, contendo informações de fácil visualização, acessadas via aplicativo instalado no celular.

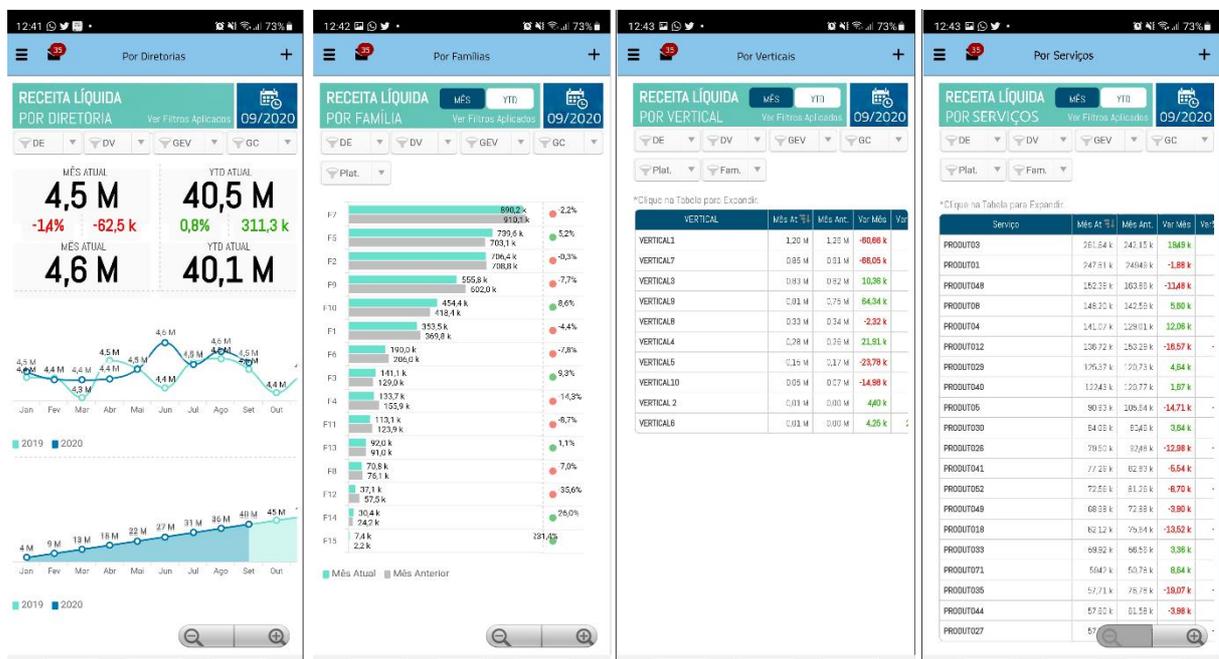


Figura 4.7 – Alguns dashboards mobile desenvolvidos.
Fonte: Elaborado pelo Autor.

4.2 – Comparação dos cenários

Após a implementação do processo de tratamento dos dados e do uso das ferramentas de BI no cotidiano da área de vendas, realizou-se uma comparação entre os cenários anterior e atual quanto às atividades impactadas pelo processo desenvolvido. As categorias avaliadas foram:

- Extração, tratamento e carga dos dados (ETL) - Antes da implementação do processo de BI, as bases de dados eram extraídas do BD pela equipe responsável e compartilhadas como planilhas Excel em pastas na rede interna da empresa. A equipe de suporte comercial filtrava as bases e distribuía para os colaboradores de suas respectivas diretorias. O tratamento das informações era limitado à bases de dados pequenas devido ao baixo poder de processamento do Excel, o que demandava a quebra de grandes arquivos em partes menores para serem trabalhados. Após a implementação, o processo de ETL foi 100% automatizado, as bases são extraídas do BD, limpas, tratadas, carregadas na nuvem e disponibilizadas para todos os usuários cadastrados, em um único processo de recorrência mensal.
- Geração de Relatórios - Antes da implementação do processo de BI, cada diretoria tratava seus dados e construía seus indicadores segundo regras próprias. Gráficos eram elaborados no Excel e as apresentações eram montadas mensalmente via Powerpoint. Para ter uma visão geral do resultado das vendas, o presidente recebia uma apresentação compilada com os indicadores de todas as diretorias. Após a implementação, todos os usuários com login têm acesso aos dashboards desenvolvidos, padronizados e atualizados de forma automática e recorrente.
- Acesso e visualização - Antes da implementação do processo de BI, os dados e relatórios eram acessados somente por computadores dentro da rede da empresa. Vários arquivos Excel precisavam ser gerados, conforme regras de visualização de dados, para que cada colaborador acessasse apenas informações de sua competência. Após a implementação, todos os usuários com login na plataforma, homologados pela equipe de TI, acessam informações por computadores, smartphones e tablets. A configuração de acesso aos dados realizada no processo de implementação permitiu que usuários distintos acessem as mesmas bases de dados, porém visualizando somente os dados de sua competência, conforme seu perfil.

- Duração do processo - A duração total do processo anterior, compreendendo as etapas de ETL, elaboração e compartilhamento de informações, era de três dias aproximadamente. O processo dependia de uma grande quantidade de colaboradores, o que aumentava a chance de erro, a burocracia e o tempo de execução. Após a implementação do processo de BI, todas as atividades são realizadas em sequência, de forma automatizada em um tempo aproximado de duas horas.

4.3 – Ganhos Obtidos

Através da comparação entre os cenários foi possível quantificar os ganhos percebidos pela área de vendas após a implementação do processo de BI. São eles:

- Automatização de 100% do processo de ETL;
- Padronização de 100% dos indicadores;
- Definição de identidade visual para os dashboards;
- Automatização de 100% do processo de notificação das atualizações dos dashboards;
- Aumento em 80% da confiabilidade na informação gerada;
- Centralização de 100% dos dados em um repositório único;
- Facilidade de acesso às bases de dados e relatórios;
- Autonomia na extração de visões personalizadas de dados;
- Possibilidade de acessar as informações por dispositivos móveis;
- Diminuição do tempo de processamento dos dados em 97%;
- Maior foco da equipe na análise das informações e na tomada de decisão.

Capítulo 5

Conclusão e Trabalhos Futuros

5.1 – Conclusão

Nos dias atuais, para ser competitiva e eficiente, uma empresa precisa possuir boas práticas de tratamento e análise de dados. As informações obtidas pela correta interpretação dos dados, quando convertida em ações, podem ser a diferença entre vida e morte para empresas na atual conjuntura de mercado. Buscando melhorar seus processos de dados e se tornarem mais competitivas, as organizações têm utilizado ferramentas de BI no tratamento dos dados e geração de seus indicadores.

No trabalho atual, foi apresentada a estruturação do processo de tratamento de dados e geração de indicadores utilizando ferramentas de BI na área de vendas de uma empresa de grande porte. Por meio de um estudo de caso, apresentou-se uma visão geral da área em foco, listando as necessidades encontradas e indicando todas as etapas existentes no processo realizado. Após a implementação, foi possível levantar os ganhos percebidos pela área em questão por meio de uma comparação entre os cenários pré e pós-implantação. Percebeu-se um grande impacto positivo na rotina de geração de relatórios da área de vendas, com ganhos expressivos em tempo de processamento, qualidade do conteúdo gerado e facilidade de acesso às informações. O fator mais expressivo percebido ao final do processo foi a mudança do foco dos colaboradores, que passou do esforço na geração dos relatórios para a interpretação mesmos e definição de estratégias.

Pode-se concluir que o correto uso das ferramentas de BI, juntamente com a adoção de rotinas eficientes no tratamento de dados e geração de indicadores, contribui de forma expressiva com os processos de inteligência de uma organização. A automatização e padronização dos processos resulta em uma grande economia de tempo que reflete na mudança dos hábitos, priorizando as ações baseadas em dados, que impactam diretamente no crescimento das empresas.

5.2 – Trabalhos Futuros

Com o objetivo de obter mais insights para a definição das estratégias, os próximos passos do processo apresentado neste trabalho relacionam-se ao desenvolvimento de algoritmos de *machine learning* para a identificação de novas oportunidades de geração de receita a partir da base de clientes da empresa, complementadas com informações externas fornecidas por empresas de dados parceiras.

Almeja-se ainda expandir o racional implementado na área de vendas para outras áreas da empresa, visando uma melhoria no processo de geração de dados para a organização como um todo. Esse processo possibilitará uma comunicação muito mais eficaz entre as áreas da empresa, tornando possível a criação de uma interface para monitoramento da saúde da organização, otimizando o processo de definição de estratégias e tomada de decisão.

Referências Bibliográficas

AGUENA, N. K. **Banco de Dados Relacional - Café com Banco de Dados**. Disponível em: <<http://www.coffeewithdatabase.com/banco-de-dados-relacional/>>. Acesso em: 17 sep. 2021.

ALIEV, R. **Migrating data with Pentaho ETL (Kettle)**. Disponível em: <<https://hybrismart.com/2017/01/15/migrating-data-with-pentaho-etl-kettle/>>. Acesso em: 28 sep. 2021.

ALOOMA TEAM. **What is Pentaho Data Integration (Kettle)?** Disponível em: <<https://www.alooma.com/answers/what-is-pentaho-data-integration>>. Acesso em: 28 sep. 2021.

ARULDOSS, M.; LAKSHMI TRAVIS, M.; PRASANNA VENKATESAN, V. A survey on recent research in business intelligence. **Journal of Enterprise Information Management**, v. 27, n. 6, p. 831–866, 7 Oct. 2014.

BARBIERI, C. Conceitos estruturantes de BI. In: **BI2: Business Intelligence**. [s.l.] Elsevier, 2011. p. 95–119.

BERTO, R.M.v.S., NAKANO, D. N. **A Produção Científica nos Anais do Encontro Nacional de Engenharia de Produção: Um Levantamento de Métodos e Tipos de Pesquisa**. *Produção*, v. 9, n. 2, p. 65-76, 2000.

BOUMAN, R.; DONGEN, J. VAN. **Pentaho Solutions: Business Intelligence and Data Warehousing with Pentaho and MySQL**. 1. ed. Indianapolis: Wiley Publishing, 2009. p. 648

CAMPOS, M. L.; ROCHA, A. V. F. Data Warehouse. In: **Congresso da Sociedade Brasileira de Computação, ed.17, 1997, Brasília**. Anais do XVII Congresso da SBC. Brasília: UNB, 1997.

CETAX. **ETL - Extract Transform Load : O que é, conceitos e definição**. Disponível em: <<https://www.cetax.com.br/blog/etl-extract-transform-load/>>. Acesso em: 19 sep. 2021.

CORTEX INTELLIGENCE. **A plataforma de inteligência para crescimento**. Disponível em: <<https://www.cortex-intelligence.com/plataforma>>. Acesso em: 21 sep. 2021.

CORONEL, C.; MORRIS, S. **Database Systems: Design, Implementation, & Management**. 13. ed. Australia: Cengage Learning, 2018. p. 816

DATABRICKS. **What is ETL?** Disponível em:

<<https://databricks.com/glossary/extract-transform-load>>. Acesso em: 19 sep. 2021.

DAYAL, U. et al. **Data integration flows for business intelligence** (M. Kersten et al., Eds.) Proceedings of the 12th International Conference on Extending Database Technology Advances in Database Technology - EDBT '09. **Anais...** In: THE 12TH INTERNATIONAL CONFERENCE. New York, New York, USA: ACM Press, 24 Mar. 2009

FERNANDES, H. M. **Banco de Dados Relacional (SQL) e Não Relacional (NoSQL) - O que são, para que servem e qual a diferença?** Disponível em: <<https://marquesfernandes.com/tecnologia/banco-de-dados-relacional-sql-e-nao-relacional-nosql-o-que-sao-para-que-servem-e-qual-a-diferenca/>>. Acesso em: 17 sep. 2021.

FIVEACTS. **Data visualization: o que é, tipos, benefícios e aplicações.** Disponível em: <<https://www.fiveacts.com.br/data-visualization/>>. Acesso em: 21 sep. 2021.

FREITAS, G. M. **Uma ferramenta de apoio à modelagem de dados dimensional.** Dissertação (Mestrado em Administração Pública) - Fundação João Pinheiro: Belo Horizonte, 2002.

FURLAN, J. D.; IVO, I. DA M.; AMARAL, F. P. **Sistema de Informação Executiva=EIS-Executive Information System: como integrar os executivos ao sistema informacional das empresas, fornecendo informações úteis e objetivas para suas necessidades estratégicas e operacionais.** São Paulo: Makron Books, 1994.

GIL, A. C. **Como Elaborar Projetos de Pesquisa.** São Paulo: Atlas, 1996.

INMON, B.; KELLEY, C. The Twelve Rules of Data Warehouse for a Client/Server World. **Data Management Review**, v. 4, n. 5, p. 6–16, May 1994.

INTELLIGENCE PARTNER. **Use Tableau to answer questions regarding your business.** Disponível em: <<https://www.intelligencepartner.com/en/use-tableau-to-answer-questions-regarding-your-business/>>. Acesso em: 17 sep. 2021.

LI, S.-T.; SHUE, L.-Y.; LEE, S.-F. Business intelligence approach to supporting strategy-making of ISP service management. **Expert systems with applications**, v. 35, n. 3, p. 739–754, Oct. 2008.

LORENA, L. H. N. **Utilização de Técnicas de Business Intelligence para avaliar o Processo Seletivo FATEC-SJC.** Monografia (Graduação em Banco de Dados) - Faculdade de Tecnologia de São José dos Campos: São José dos Campos, São Paulo, 2011.

MICROSOFT. **Visão geral do OLAP (Online Analytical Processing)**. Disponível em: <<https://support.microsoft.com/pt-br/office/vis%C3%A3o-geral-do-olap-online-analytical-processing-15d2cdde-f70b-4277-b009-ed732b75fdd6>>. Acesso em: 30 sep. 2021.

NEDIGER, M. **What is Data Visualization? (Definition, Examples, Best Practices)**. Disponível em: <<https://venngage.com/blog/data-visualization/>>. Acesso em: 21 sep. 2021.

NOLETO, C. **Banco de dados: Tipos o que é e suas diferenças!** Disponível em: <<https://blog.betrybe.com/tecnologia/bancos-de-dados/>>. Acesso em: 16 jul. 2021.

PATEL, N. **How to Use Data Visualization in Your Content to Increase Readers and Leads**. Disponível em: <<https://neilpatel.com/blog/data-visualization/>>. Acesso em: 21 sep. 2021.

PEREIRA, D. M. **Uso do Padrão OIM de Metadados no Suporte às Transformações em Ambiente Data Warehouse**. Dissertação (Mestrado em Informática) - Instituto de Matemática e Núcleo de Computação Eletrônica, Universidade Federal do Rio de Janeiro: Rio de Janeiro, 2000.

PINO, I. **Desafios e tendências da visualização de dados**. Disponível em: <<https://blog.dp6.com.br/desafios-e-tend%C3%A2ncias-da-visualiza%C3%A7%C3%A3o-de-dados-dataviz-basics-1-de-4-8208cf5fbc6b6>>. Acesso em: 21 sep. 2021.

POZZEBON, M.; FREITAS, H. M. R. Construindo um EIS (Enterprise Information System) da (e para a) empresa. **RAUSP Management Journal**, v. 31, n. 4, p. 19–30, 1996.

RAFAEL, B. **SQL, NoSQL, NewSQL: o que é, diferenças e vantagens**. Disponível em: <https://blog.geekhunter.com.br/sql-nosql-newsql-qual-banco-de-dados-usar/#E_NoSQL_O_que_e>. Acesso em: 23 sep. 2021.

SAS. **ETL: o que é e qual sua importância?** Disponível em: <https://www.sas.com/pt_br/insights/data-management/o-que-e-etl.html>. Acesso em: 18 sep. 2021.

SHARDA, R.; DELEN, D.; TURBAN, E. **Business Intelligence and Analytics: Systems for Decision Support**. 10. ed. Boston: Pearson, 2013. p. 656

SHARDA, R.; DELEN, D.; TURBAN, E. **Business Intelligence, Analytics, and Data Science: A Managerial Perspective**. 4. ed. New York, NY: Pearson, 2017. p. 512

SPEC INDIA. **What Is Pentaho Data Integration (Kettle): All There Is To Know**. Disponível em: <<https://www.spec-india.com/blog/pentaho-data-integration-kettle>>. Acesso em: 28 sep. 2021.

TABLEAU. **What is data visualization? A definition, examples, and resources.** Disponível em: <<https://www.tableau.com/pt-br/learn/articles/data-visualization>>. Acesso em: 21 sep. 2021.

TECHLIB. **Definição de RDBMS.** Disponível em: <<https://techlib.wiki/definition/rdbms.html>>. Acesso em: 17 sep. 2021.

TEJADA, Z. **Dados não relacionais e NoSQL - Azure Architecture Center | Microsoft Docs.** Disponível em: <<https://docs.microsoft.com/pt-br/azure/architecture/data-guide/big-data/non-relational-data>>. Acesso em: 17 sep. 2021.

WILLIAMS, A. **Column-Oriented Databases, Explained - KDnuggets.** Disponível em: <<https://www.kdnuggets.com/2021/02/understanding-nosql-database-types-column-oriented-databases.html>>. Acesso em: 17 sep. 2021.

YIN, R. K. **Estudo de Caso – Planejamento e Método.** 2. ed. São Paulo: Bookman, 2001.