UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Instituto de Ciência e Tecnologia
Câmpus de Sorocaba

JOSÉ MENEZES DE OLIVEIRA JUNIOR

Automatização de processo de contas a receber através da associação de ETL (*Extract, Transform, Load*) e Power BI

Sorocaba
2022

JOSÉ MENEZES DE OLIVEIRA JUNIOR

AUTOMATIZAÇÃO DE PROCESSO DE CONTAS A RECEBER ATRAVÉS DA ASSOCIAÇÃO DE ETL (*EXTRACT, TRANSFORM, LOAD*) E POWER BI

Trabalho de Conclusão de Curso apresentado ao Instituto de Ciência e Tecnologia de Sorocaba, Universidade Estadual Paulista (UNESP), como parte dos requisitos para obtenção do grau de Bacharel em Engenharia de Controle e Automação.

Orientador(es): Prof. Dr. Márcio Alexandre Marques.

Sorocaba

2022

O48a Oliveira Junior, Jose Menezes de
Automatização de processo de contas a receber através da associação de ETL (extract, transform, load) e Power BI / Jose Menezes de Oliveira Junior. -- Sorocaba, 2022
57 p.

Trabalho de conclusão de curso (Bacharelado - Engenharia de Controle e Automação) - Universidade Estadual Paulista (Unesp), Instituto de Ciência e Tecnologia, Sorocaba
Orientador: Marcio Alexandre Marques

1. Automação por software. 2. Business Intelligence. 3. Power BI. 4. Pentaho Data Integration. 5. PostgreSQL. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do Instituto de Ciência e Tecnologia, Sorocaba. Dados fornecidos pelo autor(a).

UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Instituto de Ciência e Tecnologia
Câmpus de Sorocaba

AUTOMATIZAÇÃO DE PROCESSOS DE CONTAS A RECEBER ATRAVÉS DA ASSOCIAÇÃO DE ETL (*EXTRACT, TRANSFORM, LOAD*) E POWER BI

JOSÉ MENEZES DE OLIVEIRA JUNIOR

ESTE TRABALHO DE GRADUAÇÃO FOI JULGADO ADEQUADO COMO PARTE DO REQUISITO PARA OBTENÇÃO DO GRAU DE **BACHAREL EM ENGENHARIA DE CONTROLE E AUTOMAÇÃO**

Prof. Dr. Everson Martins
Coordenador

BANCA EXAMINADORA:

Prof. Dr. Márcio Alexandre Marques

Prof. Dr. Everson Martins

Prof. Dra. Marilza Antunes Lemos

Dezembro de 2022

DEDICATÓRIA

dedico esse trabalho à minha esposa, Rafaela Giardini Menezes, que todos os dias faz o necessário para me tornar um ser humano melhor e à minha família, base fundamental na construção da minha vida e carreira

AGRADECIMENTOS

À UNESP, por toda estrutura e acolhimento durante os anos de graduação, grande responsável pela minha formação cultural e acadêmica e que é responsável pelo meu sucesso profissional.

Ao Prof. Dr. Márcio Alexandre Marques, principal docente na minha formação dentro da universidade, sempre justo e de enorme suporte a mim e aos meus colegas.

JUNIOR, J. M. O. **Automatização de processo de contas a receber através da associação de ETL (Extract, Transform, Load) e Power BI.** 2022 (Bacharel em Engenharia de Controle e Automação) – Instituto de Ciência e Tecnologia de Sorocaba – ICTS – Universidade Estadual Paulista, 2022

RESUMO

Este trabalho aborda a utilização de softwares gratuitos para automação de processos e análise de dados de uma empresa privada. O objetivo é realizar a automação de um processo manual de contas a receber na área financeira dessa empresa, pois o responsável por esse processo gasta cerca de 70h por mês para realizá-lo. Foram utilizadas as ferramentas Pentaho Data Integration, Power BI e um banco de dados PostgreSQL, escolhidas por serem gratuitas e de fácil manipulação. O *dashboard* construído em Power BI, além da versão local, também foi disponibilizado na plataforma *online* da solução, possibilitando que outras pessoas da empresa possam acessar a ferramenta e os resultados gerados. Com a implementação dessa aplicação, foi possível agilizar a conferência dos itens vendidos pela empresa e com isso, a solução proposta reduziu em aproximadamente 80% o tempo na execução da tarefa. Além disso, a ferramenta permitiu a criação de novos indicadores que trouxeram mais insumos sobre clientes e fornecedores que se destacam, bem como clientes que mais devem para a empresa.

Palavras-chave: Automação por software; *Business Intelligence*; Power BI; Pentaho Data Integration; PostgreSQL; *Dashboard*

JUNIOR, J. M. O. **Accounts Receivable process automation using ETL (Extract, Transform, Load) and Power BI.** 2022 (Bachelor's degree Automation and Control Engineering) – Science and Technology Institute – Sorocaba, UNESP 2022.

ABSTRACT

This paper discusses the use of free software for process automation and data analysis in a private company. The objective is the automation of a manual process of accounts receivable performed in the financial area of this company, because the person responsible for this process spends about 70 hours per month to perform it. Were used Pentaho Data Integration, Power BI and a PostgreSQL database, chosen for being free and for their easy utilization. The dashboard built in Power BI, besides the local version, was also made available on the solution's online platform, allowing other people in the company to access the tool and the results generated. With the implementation of this application, it was possible to speed up the checking of items sold by the company, and with this, the proposed solution reduced by approximately 80% the time in performing the task. In addition, the tool allowed the creation of new indicators that brought more inputs about customers and suppliers that stand out, as well as customers who owe more to the company.

Keywords: Software automation; *Business Intelligence*; Power BI; Pentaho Data Integration; PostgreSQL; *Dashboard*

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

BD	Banco de Dados
BI	*Business Intelligence* – Inteligência de Negócio
CLP	Controlador Lógico Programável
CSV	*Comma-Separated Values* – Valores Separados por Vírgula
DBMS	*Database Management System* – Sistema Gerenciador de Banco de Dados
DSN	*Data Source Name* – Nome da Fonte de Dados
ETL	*Extract, transform, load* – Extração, transformação e carregamento
JSON	*JavaScript Object Notation* – Notação p/ Objeto JavaScript
KPI	*Key Performance Indicator* – Indicador Chave de Performance
NoSQL	*Not Structured Query Language* – Não Linguagem de Consulta Estruturada
OCR	*Optical Character Recognition* – Reconhecimento Ótico de Caracteres
OLAP	*Online Analytical Processing* – Processamento Analítico Online
PBI	Power BI
PDI	Pentaho Data Integration
RegEx	*Regular Expression* – Expressão Regular
RPA	*Robotic Process Automation* – Automação de Processos com Robôs
SQL	Linguagem de Consulta Estruturada
TI	Tecnologia da Informação
YTD	*Year to Date* – Do começo do ano até o momento

SUMÁRIO

1. INTRODUÇÃO

O avanço da tecnologia é tido por muitos estudiosos como a alavanca de desenvolvimento dos países: quanto mais tecnológico é um país, mais este se torna competitivo dentro da economia mundial. Essa análise se estende para as grandes e pequenas empresas e indústrias dentro dos países: para que essas possam prosperar e manter seu espaço na economia, a corrida pela substituição de processos manuais e ultrapassados por processos automáticos se torna cada vez maior.

Dentro dessa vertente, o avanço das ferramentas computacionais que minimizam o tempo de trabalho do homem está em constante crescimento – não é de hoje que o homem procura transferir o tempo gasto com processos repetitivos para processos que de fato gerem maior valor [1]. Assim, os processos automatizados por robôs (RPA - *Robotic Process Automation*) e o ETL - Extração, Transformação e Carregamento (*Extract, Transform, Load*) são cada vez mais utilizados nas empresas para permitir que trabalhadores gastem seu tempo em atividades intelectuais que não podem ser feitas pelo computador, enquanto os processos repetitivos sejam deixados com ele [2].

Quando pensamos em automação de processos temos dois principais caminhos: a automação industrial, por meio de CLP, robôs e esteiras e a automação por software, pouco abordada durante a graduação, que permite através de ferramentas específicas realizar processos repetitivos como a conferência de documentos, reconhecimento ótico de caracteres (OCR), envio automático de *e-mails* e consolidação de arquivos em diferentes formatos em apenas um.

Com as tarefas repetitivas em modo automático, temos mais tempo para os colaboradores e gestores poderem analisar os dados minerados e é nesse momento que entra a importância das ferramentas de *Business Intelligence*, como o Power BI. Essas ferramentas são capazes de agrupar, organizar, analisar e disponibilizar os dados de forma agradável utilizando o visual, permitindo uma melhor reflexão sobre pontos importantes para a construção das estratégias que serão adotadas [3].

Com a junção de técnicas de automação por software, análise e apresentação de dados, temos uma área já consolidada mundialmente mas ainda com enorme potencial para os engenheiros de controle e automação atuarem, permitindo um grande desenvolvimento profissional em seu leque de aplicação: poder atuar de ponta a ponta dentro de grandes corporações ou em setores públicos com a substituição no modo manual de gerar dados até sua distribuição/apresentação em painéis que possibilitem a ótima utilização desses dados.

Dessa maneira, este trabalho de graduação tem como motivação, apresentar uma solução de automação de um processo empresarial ao transformar um método extremamente repetitivo

de contas a receber realizado em uma empresa privada através da conferência de extratos pagos pelo cliente. Atualmente, o analista financeiro leva 70h por mês para poder conferir os mais de 2000 itens vendidos – comparando cada extrato pago pelos diversos clientes com a base de vendas da sua empresa. Para isso, o trabalho tem como objetivos:

1) Estudar e apresentar conceitos sobre ETL (*Extract, Transform, Load*), Banco de Dados e ferramentas de visualização de dados

2) Utilizar as seguintes tecnologias para automatização do processo:

   a. Banco de Dados SQL

   b. Ferramenta ETL (Pentaho)

   c. Visualização de Dados (Power BI)

3) Analisar os dados da empresa para elaborar o processo de Contas a Receber automático

4) Desenvolver painel de Contas a Receber através do Power BI para ajudar na visualização dos dados

2. REVISÃO BIBLIOGRÁFICA

2.1 Banco de Dados

Um banco de dados é uma coleção organizada de informações – ou dados – estruturadas, normalmente armazenadas eletronicamente em um sistema de computador. Um banco de dados é geralmente controlado por um sistema de gerenciamento de banco de dados (DMBS – *Data Base Management System*). Juntos, os dados e o DBMS, em conjunto com os aplicativos associados a eles, são chamados de sistema de banco de dados, comumente abreviados para apenas banco de dados (BD). Os dados nos tipos mais comuns de bancos de dados em operação atualmente são modelados em linhas e colunas em uma série de tabelas para tornar o processamento e a consulta de dados mais eficiente. Os dados podem ser facilmente acessados, gerenciados, modificados, atualizados, controlados e organizados. A maioria dos bancos de dados usa a linguagem de consulta estruturada (SQL) para escrever e consultar dados [4].

O Sistema Gerenciador de Banco de Dados disponibiliza recursos para definir, construir, manipular, compartilhar, proteger e manter bancos de dados. Neste sentido, existem diferentes soluções que permitem o gerenciamento de banco de dados, sendo as principais Oracle, SQL Server, PostgreSQL, MySQL, AWS e Access. Os sistemas de gerenciamento são, portanto, programados de modo a permitir o acesso, extração e manipulação por diferentes softwares de terceiros. Cria-se assim uma interface entre os usuários e a informação [5].

Por exemplo, no momento em que compramos algo com cartões de débito ou crédito ou então, comprando produtos de forma *online*, utilizando os mais diversos tipos de *e-commerce*, estamos fazendo interações com pelo menos um banco de dados [6].

Os bancos de dados se tornaram base para criação de diversos tipos de softwares e aplicações, tanto no setor público quanto no privado, o que foi possibilitado devido a agilidade na criação dessas ferramentas tanto *offline* quanto *online*, com um ótimo custo-benefício [7].

Os bancos de dados podem ser divididos principalmente em Bancos de Dados Relacionais e Não-Relacionais.

2.1.1 Bancos de Dados Relacionais

O surgimento dos bancos de dados relacionais, segundo Wade e Chamberlain [6], pode ser considerado como uma das histórias mais bem sucedidas e de maior impacto na vida dos seres humanos. No início dos anos 1970, Edgar Frank Codd trabalhando no escritório da IBM® escreveu um artigo que sugeria um novo modo para armazenar informações em um computador, chamado de "*A Relational Model of Data for Large Shared Databanks",* que fora mais tarde batizado como "Modelo de Dados Relacional".

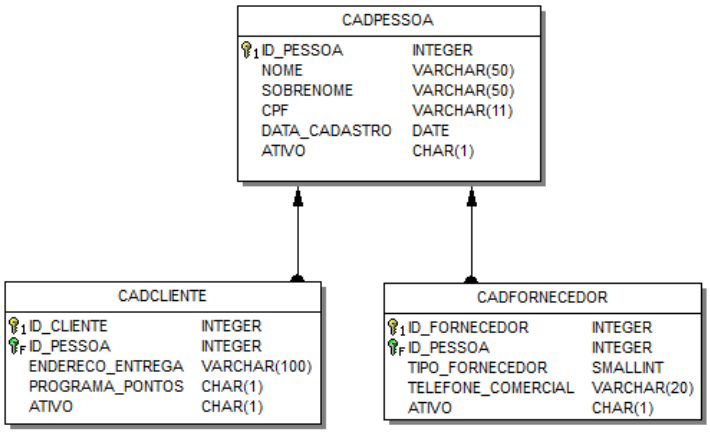
Ainda segundo Wade e Chamberlain, houve uma demora de aproximadamente 10 anos para aceitarem de fato e começarem a utilizar em sistemas comerciais sua proposta. Esse modelo proporcionou aumento na produtividade dos desenvolvedores que passaram a optar pelas estruturas de dados em linhas e colunas, organizados por tabelas ao invés de índices de conteúdo [6].

Os bancos de dados relacionais são um tipo de banco de dados que armazena e provê acesso a dados que estão relacionados uns aos outros. São baseados no modelo de relacionamento. Nesse tipo de BD, cada linha da tabela é um registro com um ID único, chamado de chave. As colunas da tabela comportam os atributos dos dados e cada registro normalmente tem um valor para cada atributo, facilitando assim o estabelecimento de relacionamentos entre os dados [8].

O modelo de relacionamento traz consigo que as estruturas de dados lógicas estão separadas das estruturas de armazenamento físicas. Essa separação se traduz no fato dos administradores poderem manusear a estrutura física sem afetar o que está contido dentro dela – por exemplo, ao trocarmos o nome de um banco de dados em si, as tabelas contidas dentro dele não serão renomeadas. Essa separação também se aplica nas operações do banco, onde as operações lógicas permitem que uma aplicação especifique qual o conteúdo que ela necessita e as operações físicas determinam como os dados serão acessados [8].

A Figura 1 apresenta um breve exemplo de tabelas conectadas em uma estrutura de banco de dados relacional.

Figura 1: Exemplo da estrutura de dados relacionais



Fonte: [9]

Podemos verificar na figura 1, a existência de três diferentes tabelas: CADPESSOA, onde serão armazenadas as informações e registros de diferentes pessoas no banco; CADCLIENTE, onde serão cadastradas as informações relacionadas a clientes; e CADFORNECEDOR, onde serão cadastradas as informações relacionadas aos fornecedores. Nelas, o relacionamento é feito

através da chave ID_PESSOA, onde garantimos que cada registro de cliente ou fornecedor esteja conectado a base de cadastro de pessoas.

2.1.2 Bancos de Dados Não Relacionais

Um banco de dados não relacional é um banco de dados que não usa o esquema de tabela de linhas e colunas encontrado na maioria dos sistemas de banco de dados tradicionais. Em vez disso, os bancos de dados não relacionais usam um modelo de armazenamento otimizado para os requisitos específicos do tipo de dados que está sendo armazenado. Por exemplo, os dados podem ser armazenados como pares chave/valor simples, como documentos JSON ou como um gráfico que consiste em bordas e vértices [10].

Também são chamados de NoSQL, ou seja, bancos que não usam o sistema SQL para realizar as consultas e sim outras linguagens de programação para consultar os dados. Por não usarem o modelo relacional padrão de banco de dados, são mais específicos na tipagem de dados que serão armazenados e na forma de consulta. Por esse motivo, não são bem quistos para o gerenciamento de dados transacionais como os modelos relacionais. Exemplos disso são os armazenamentos de dados de documentos, colunas, chave/valor e gráficos que serão brevemente explicados a seguir:

a. **Dados de Documentos:** esse tipo de armazenamento gerencia um conjunto de campos de cadeia de caracteres nomeadas e valores de dados de objeto em uma entidade conhecida como documento. Esses repositórios utilizam o formato JSON para armazenar os dados. Os dados nos campos de um documento podem ser de vários tipos (escalar, listas ou dicionários) e são expostos a um sistema de gerenciamento de armazenamento, permitindo que um aplicativo consulte e filtre dados utilizando os valores nesses campos. A Figura 2 apresenta um exemplo de estrutura desse modelo de armazenamento.

Figura 2: Exemplo da estrutura do armazenamento de dados de documento

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

Fonte: [10]

b. **Dados de Colunas:** nesse modelo, os dados são organizados em colunas e linhas. Embora se pareça conceitualmente com um banco de dados relacional, se diferencia na forma desnormalizada de estruturação dos seus dados: as colunas são divididas em grupos, onde cada um desses grupos (chamados de família de colunas) está logicamente relacionado e normalmente não são manipulados de forma separada. Dentro de cada família de colunas, novas colunas podem ser adicionadas dinamicamente e as linhas não necessitam ter a mesma organização de valores. A Figura 3 apresenta duas estruturas que exemplificam o modelo, onde temos em linhas diferentes, colunas diferentes que não seguem um padrão. Nesse exemplo, com duas famílias de colunas (*Identity* e *Contact Info)*, percebemos que existe uma variação no número de colunas para cada uma das linhas. A chave para consulta de cada uma das linhas é dada pelo índice presente no *CustomerID*.

Figura 3: Exemplo da estrutura do armazenamento de dados de colunas

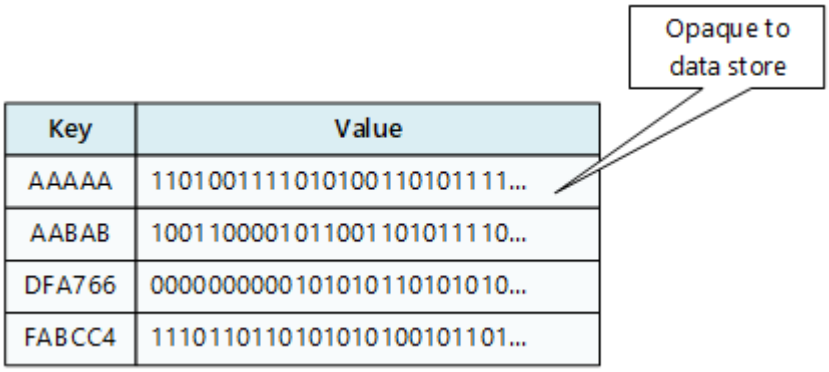
CustomerID	Column Family: Identity
001	**First name:** Mu Bae **Last name:** Min
002	**First name:** Francisco **Last name:** Vila Nova **Suffix:** Jr.
003	**First name:** Lena **Last name:** Adamcyz **Title:** Dr.

CustomerID	Column Family: Contact Info
001	**Phone number:** 555-0100 **Email:** someone@example.com
002	**Email:** vilanova@contoso.com
003	**Phone number:** 555-0120

Fonte: [10]

c. **Dados de Chave/Valor:** nesse armazenamento os dados são associados a uma chave exclusiva. Nesse modelo, temos suporte apenas a operações de exclusão, inserção e consultas, ou seja, caso o usuário queira modificar um valor, deverá substituir todos os dados presentes na chave. Esse modelo é mais indicado para aplicações que usem pesquisas simples pelo tempo de consulta, não sendo então ideal para casos onde as consultas devem ser feitas em várias tabelas para união de dados. A Figura 4 exemplifica essa estrutura.

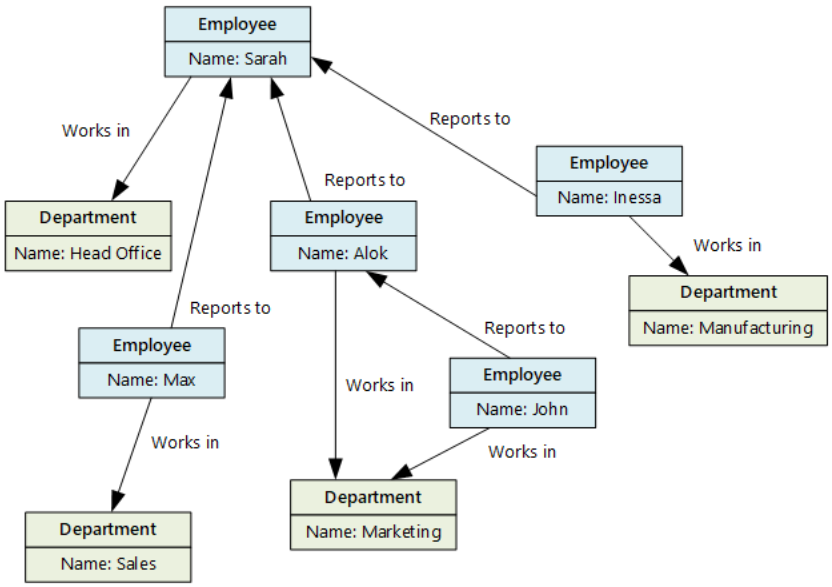
Figura 4: Exemplo da estrutura do armazenamento por chave/valor



Fonte: [10]

d. **Dados de Gráficos:** essa estrutura armazena dois tipos de informações: nós e bordas. Ambos podem ter propriedades sobre outros nós e bordas, semelhante a colunas em tabelas. O objetivo dessa estrutura é permitir que um aplicativo consulte com eficácia as relações entre os dados e suas entidades. A Figura 5 exemplifica dados sobre uma organização, onde as entidades são funcionários e departamentos e as bordas indicam os relacionamentos de relatórios e o departamento que esses funcionários trabalham.

Figura 5: Exemplo da estrutura do armazenamento de dados de gráficos



Fonte: [10]

2.2 ETL - *Extraction, Transform and Load*

Com o grande aumento no volume de dados causado pela rápida expansão da tecnologia e meios de comunicação desde meados dos anos 1990-2000, empresas e instituições públicas cada vez mais lidam com diferentes fontes de dados. Contudo, por mais que estejam em fontes diferentes, muitos desses dados tratam do mesmo assunto ou então, podem ser conectados para gerar diversas análises. Nesse momento, entra a importância das ferramentas de ETL.

O termo ETL vem de *Extraction, Transform and Load* - extração, transformação e carregamento. É um processo que habilita a "limpeza", normalização e tratamento de diferentes fontes de dados para "alimentar" um único local para análises ou suporte à alguma operação.

No mundo da computação, em geral, o ETL é utilizado para integração de múltiplas fontes ou aplicações - possivelmente até de diferentes domínios – visando ajustar e selecionar os dados corretamente para adequação à operação em três diferentes estágios [11]:

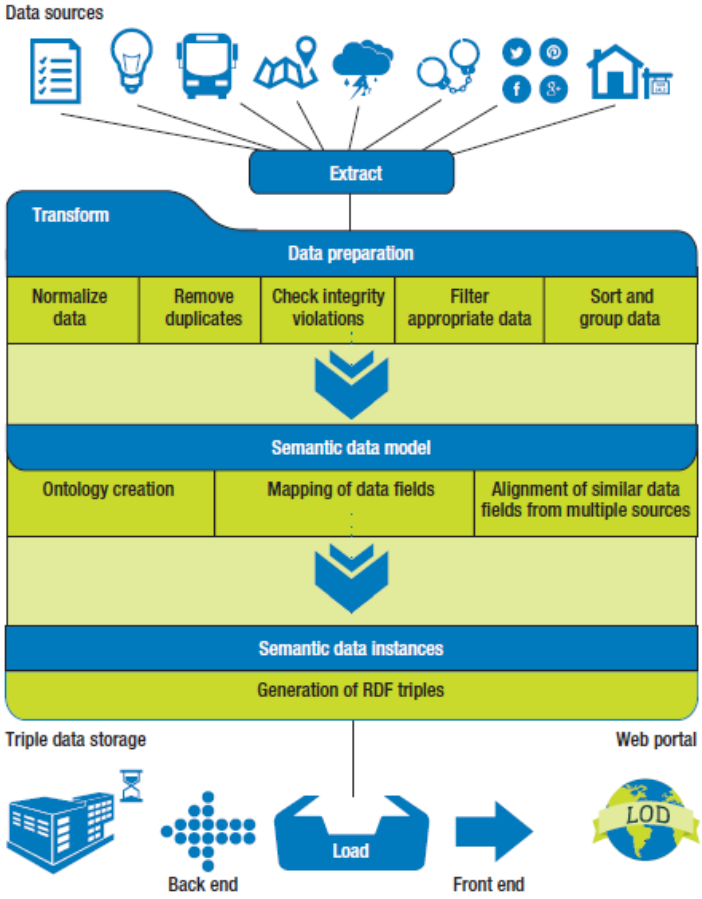
a. **Extração:** envolve a aquisição de dados das fontes apropriadas, onde o dado normalmente está disponível em formatos como CSV (*comma-separated values*) valores separados por vírgula, Excel, .txt ou através de APIs onde o resultado normalmente vem por JSON.
b. **Transformação:** envolve todo o tratamento do dado, como limpeza, adequação, formatação e mudanças de tipo (texto, inteiro, *booleano* etc.) para adequação à estrutura que devemos construir. Transformações típicas envolvem normalização dos dados, remoção de duplicatas, checagem por violação de integridade das limitações do banco,

filtros por RegEx (*regular expression*) expressão regular, muito utilizada para facilmente detectar padrões dentro de textos para correções, *sorting*, agrupamento e claro, aplicação de qualquer outra função específica.

c. **Carregamento:** envolve a propagação do dado para um banco de dados alvo, *data mart* ou *data warehouse* para consumo do usuário.

A Figura 6 ilustra o *framework* do processo durante essa integração dos dados.

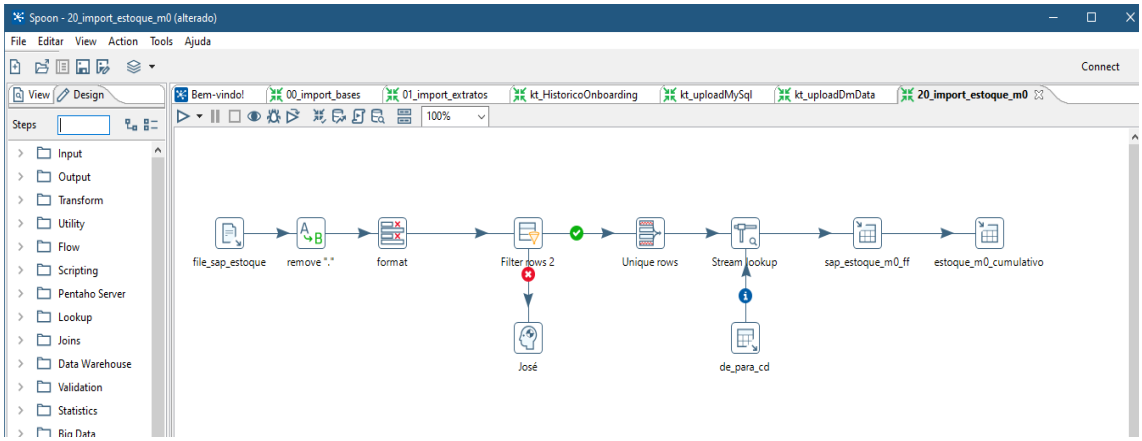
Figura 6: *Framework* de integração de dados por ETL



Fonte: [11]

Atualmente existem diversas ferramentas de ETL como o Azure Data Factory, Databricks, Amazon Glue, Amazon Lambda, Informatica Power Center, Talend, Pentaho Kettle (que iremos utilizar nesse trabalho) entre vários outros. Algumas ferramentas, como o Azure Data Factory, o Talend e o Pentaho Kettle têm uma interface amigável, onde para a realização do ETL é necessário apenas arrastar e configurar blocos, que irão desenvolver o código de programação por trás. A Figura 7 apresenta um trecho de um processo ETL desenvolvido em Pentaho Kettle.

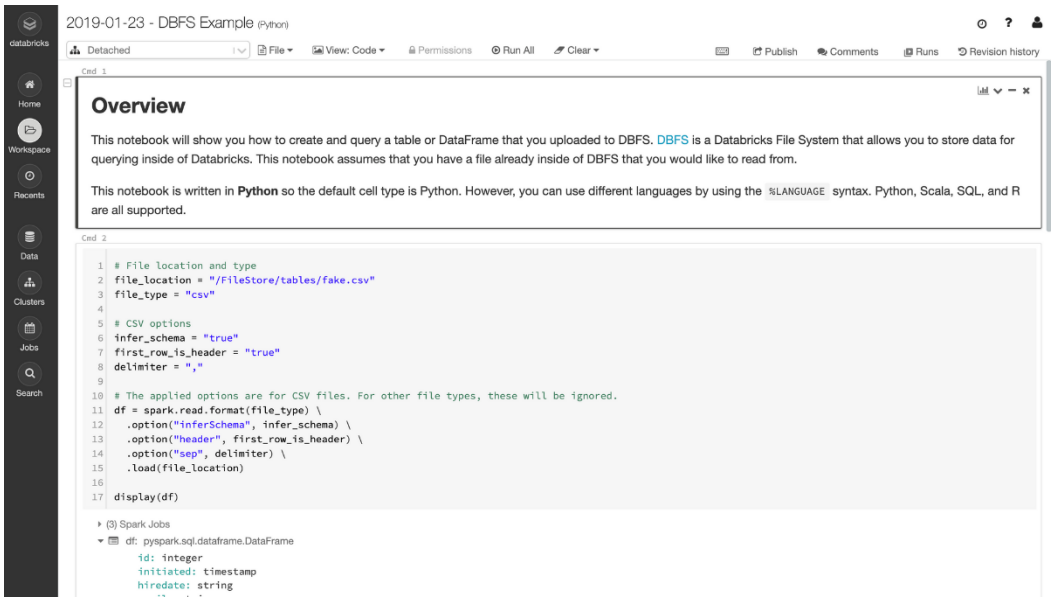
Figura 7: Exemplo de processo ETL no Pentaho Kettle.



Fonte: Autoria própria.

Já outros serviços, como o Databricks, possibilitam o processo ETL através de uma IDE comum, onde o usuário pode programar em Python, Scala e R e utilizar as diversas bibliotecas já existentes para poder desenvolver. Nesse caso, o próprio sistema da Databricks já faz todo o arquivamento e gerenciamento dos repositórios e possibilita programação simultânea, inserção de comentários e títulos nas células de comando. A Figura 8 apresenta um exemplo de *notebook* no Databricks, utilizando Python.

Figura 8: Interface do Databricks com pequeno trecho de código.



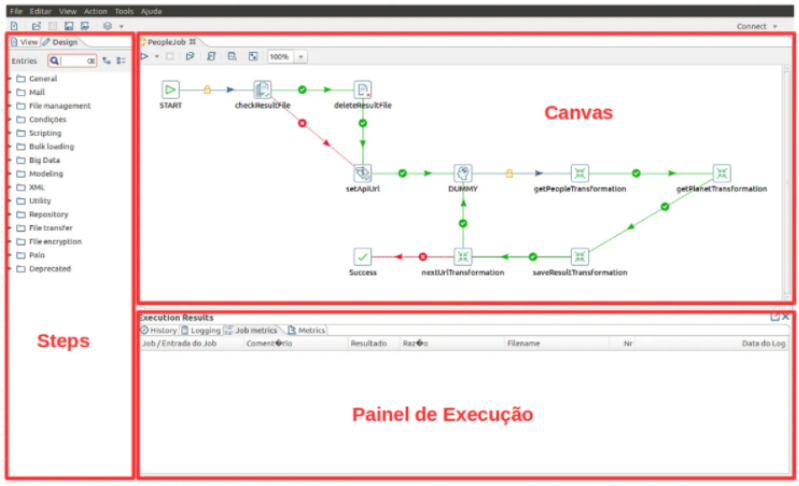
Fonte: [12]

2.2.1 Pentaho Data Integration

O Pentaho Data Integration (PDI) é um software gratuito pertencente à suíte Pentaho, a qual é composta por outros softwares voltados para a construção de aplicações/soluções de *business intelligence*, ou seja, programas que fazem a extração dos dados de sistemas de origem,

tratamento e preparação, e carregamento em outros sistemas de destino ou componentes da suíte para análise, estudo ou apenas consulta pelo usuário final. No PDI é possível criar processos de ETL para alimentação de bancos de dados e/ou criação de arquivos de texto, CSV e Excel. A Figura 9 exemplifica a interface gráfica do Pentaho Data Integration, com as três principais paletas da ferramenta.

Figura 9: Interface gráfica do Pentaho Data Integration



Fonte: [13]

O PDI é separado em três ambientes: *Spoon*, *Kitchen* e *Pan*. O *Spoon* é a interface gráfica, onde a lógica de ETL é construída e é caracterizado por dois principais arquivos, os Jobs e as Transformações. A interface, como apresentada na Figura 9, é dividida em três principais áreas:

1) *Steps*: paleta no canto esquerdo do software, onde os processos/blocos estão separados por categoria para facilitar a utilização pelo usuário
2) *Canvas*: folha em branco onde os processos da paleta *Steps* são "soltos" para criação da lógica ETL. Cada "folha" tem início e fim, com os blocos sendo programados para cumprirem determinada função. Aqui é onde as Transformações e Jobs são montados:
    a. As Transformações são processos com vários blocos de diferentes funções para cumprir o ETL, onde o passo a passo desde como a extração/leitura dos dados é realizada até o seu carregamento no destino final. Ela normalmente contém a leitura de dados de tabelas em bancos de dados ou Excel e CSV; seleção de campos específicos; concatenação de valores de campos distintos; matemática sendo aplicada (soma,

divisão, multiplicação) em valores da tabela; cruzamento de dados entre tabelas diferentes; aplicação de expressões regulares para limpeza.

b. Os Jobs são os processos que organizam e sequenciam a execução das transformações, para que elas aconteçam em uma determinada ordem para cumprir a função da solução desenvolvida.

3) Painel de Execução: o painel de execução é onde o *debug* do ETL acontece, sendo assim possível ler os *logs* de execução, as métricas de execução (tempo, volume de dados, número de linhas lidas e escritas etc.) e também o histórico de execução das Transformações e Jobs.

Já o *Pan* é o programa que opera por linha de comando (ou seja, sem interface gráfica) e é responsável pela execução das transformações quando estas estão finalizadas e já em produção (funcionamento sem intervenção e supervisão humana), onde as transformações que serão executadas devem ter sido agendadas em intervalos específicos do dia. Por fim, de forma paralela ao *Pan*, o *Kitchen* executa os Jobs que foram agendados previamente em intervalos regulares.

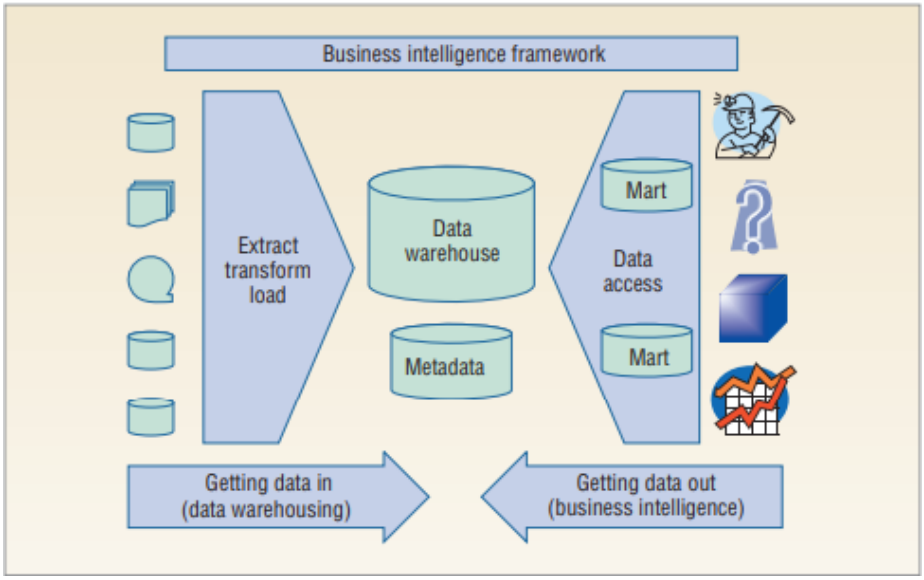
2.3 *Business Intelligence* (BI)

No começo dos anos 1970 foram desenhados os primeiros sistemas de suporte à decisão. Eles eram um contraste para os sistemas que apenas processavam transação ou aplicativos operacionais como os de entrada de pedido, controle de estoque e sistemas de pagamento [14].

Com o passar dos anos, várias aplicações para suporte a decisão – informações executivas, processamento *online* analítico (OLAP) e análises preditivas – surgiram e expandiram o domínio dessa área de suporte à decisões. No começo dos anos 1990 Howard Dressner, que na época era um analista no grupo Gartner, cunhou o termo *business intelligence*. Atualmente BI é um termo extremamente popularizado para descrever aplicações analíticas [14].

A Figura 10 apresentar um *framework* de *business intelligence* feito por Watson e Wixom [14]. Importante destacar a presença do processo ETL, no início do *framework*.

Figura 10: *Framework* de *Business Intelligence*



Fonte: [14].

As ferramentas de *business intelligence* (BI) são tipos de software de aplicativo que coletam e processam grandes quantidades de dados não estruturados de sistemas internos e externos, incluindo livros, jornais, documentos, registros médicos, imagens, arquivos, *emails*, vídeos e outras fontes comerciais. Estas ferramentas auxiliam na preparação de dados para análises, possibilitando a criação de relatórios, painéis e visualizações de dados. Os resultados dão aos funcionários e gerentes o poder de acelerar e aprimorar as tomadas de decisões, aumentar a eficiência operacional, localizar potenciais de receita, identificar as tendências do mercado, apresentar KPIs genuínos e apontar novas oportunidades de negócios [15].

Normalmente utilizadas para consultas e relatórios mais simples diretos de dados comerciais, as ferramentas de business intelligence podem combinar um vasto conjunto de aplicativos de análise de dados, incluindo consultas e análises ad hoc, relatórios empresariais, processamento analítico online (OLAP), BI móvel, BI em tempo real, BI operacional, nuvem e software como BI de serviço, BI de software livre, BI colaborativo e inteligência de localização. Elas também podem incluir software de visualização de dados para a criação de gráficos, bem como ferramentas para criação de painéis de BI e tabelas de desempenho que exibem as métricas e KPIs do negócio para proporcionar dados essenciais à empresa de uma forma simples [15].

O crescimento operacional e estratégico das grandes companhias certamente foi influenciado pelas soluções de BI. O grupo Gartner, consultoria de grande renome mundial na área de serviços estratégicos e tecnológicos, elabora anualmente um quadrante mágico das

plataformas de BI onde apresenta quais são os expoentes dessa categoria. Nesse quadrante, a empresa separa as plataformas como Líderes, Desafiantes, Visionários e Empresas “Nichadas” com base em critérios selecionados. A Figura 11 apresenta o quadrante da consultoria.

Figura 11: Quadrante Mágico da Gartner Group

Fonte: [16]

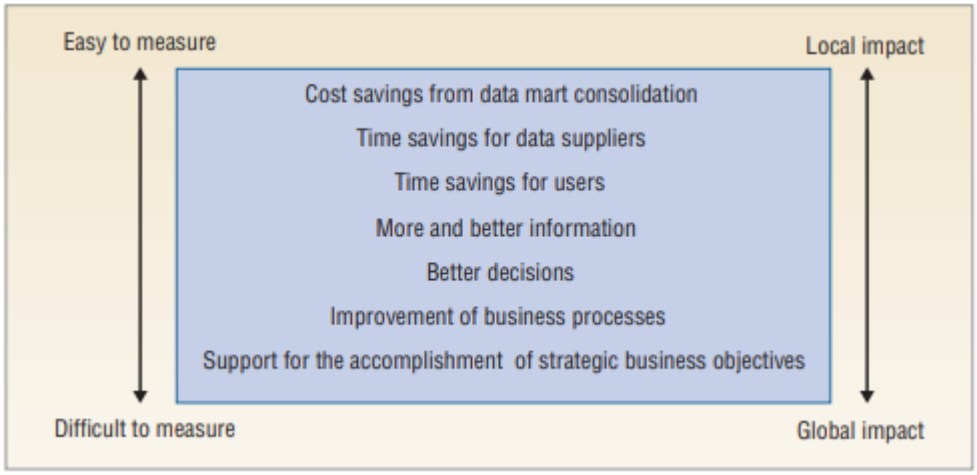
Dentro dessas plataformas, destaque para o Power BI - Microsoft (que foi utilizado nesse trabalho) que pelo 3º ano consecutivo lidera o *ranking*. Além dele, Tableau e Qlik compõe as três primeiras posições como as líderes de mercado.

O Quadrante Mágico da Gartner enquadra as plataformas nas quatro categorias citadas anteriormente conforme sua capacidade de execução de análises – em termos de recursos analíticos, facilidade de execução, velocidade, conectividade com outras plataformas - e

completude das visões que são possíveis de gerar, ou seja, quão integráveis são seus conteúdos e como o compartilhamento/publicação dessas visões é feita.

As soluções de BI oferecem *insights* valiosos sobre o que está acontecendo com as organizações no momento pela sua eficiência e com o passar do tempo, ao invés de apenas se questionarem sobre o que está acontecendo no momento. Assim, as organizações expandem sua gama de análises para entender o porquê aquilo está acontecendo e até mesmo o que poderá acontecer em decorrência do retrato atual apresentado pelos dados. Os níveis de benefício, em um primeiro momento, são fáceis de mensurar mas com um impacto local. Com a maior utilização e suporte da ferramenta, eles passam a ter um impacto global e se tornam difíceis de mensurar pelos benefícios trazidos [14]. A Figura 12 ilustra essa ideia.

Figura 12: Benefícios da aplicação do BI



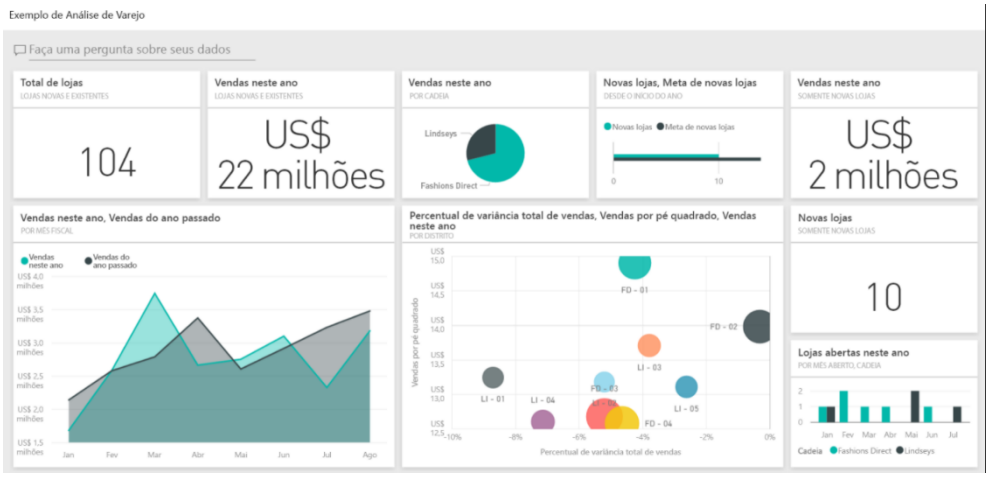
Fonte: [14]

2.3.1 Power BI

Lançado ao público pela Microsoft® em 2015, o Power BI é o software líder de mercado – como apresentado na Figura 11, em análise pela Gartner – responsável por entregar análise, modelagem e visualização de dados em um único pacote. É muito utilizado em análise de negócios devido a sua versatilidade em acoplar bases de dados de diversos tipos, apresentação e aplicações gráficas que permitem análises sensíveis, e atualização de dados em tempo real. Possui interface amigável e permite que qualquer pessoa, independente do seu grau de familiaridade com esse tipo de ferramenta, elabore gráficos simples para auxiliar em suas rotinas – bastando apenas conhecer um pouco sobre computação [17].

Com a ferramenta, o usuário tem a capacidade de criar diferentes tipos de *dashboards* que podem ser atualizados em tempo real ou com agendamentos. A Figura 13 apresenta um exemplo de *dashboard* criado em Power BI.

Figura 13: Exemplo de *dashboard* feito em Power BI

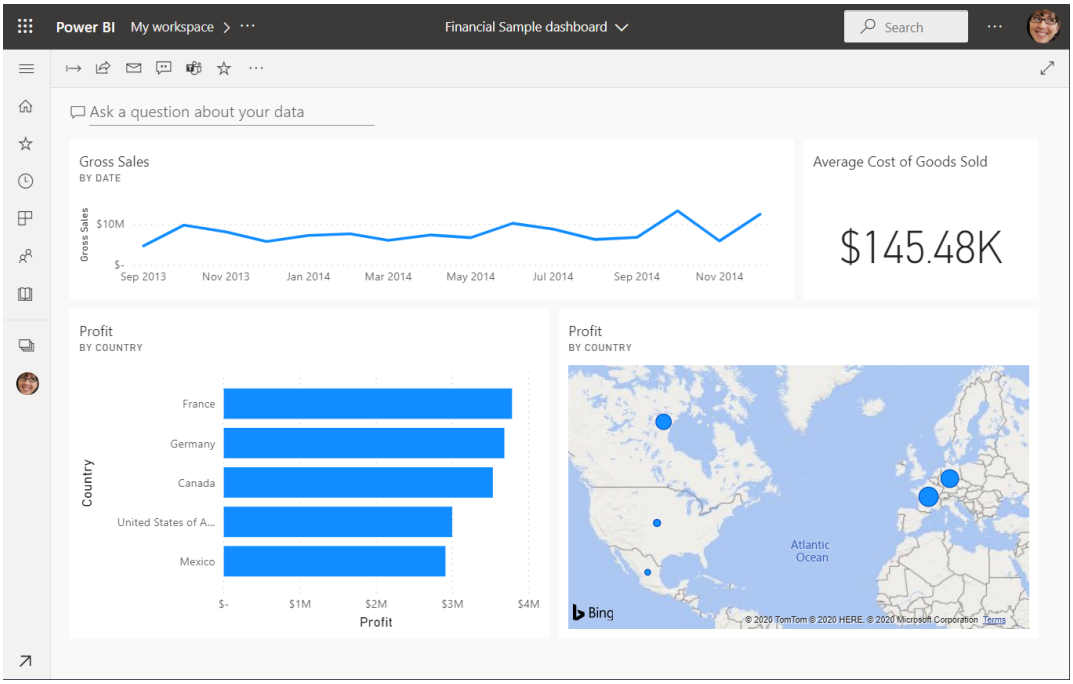


Fonte: [18]

Por se tratar de uma ferramenta de fácil manuseio, as áreas de negócio das empresas se tornam autosuficientes na geração e manutenção dos seus relatórios analíticos, levando a tomada de decisões mais rápidas e efetivas.

Todos os *dashboards* criados localmente na máquina do usuário podem ser publicados através do Power BI Service, onde outros usuários da companhia, além do criador, podem acessar as informações na nuvem e com total capacidade de manipular, selecionar e filtrar os dados que deseja. A Figura 14 apresenta um exemplo da visualização de *dashboard* publicado na nuvem [19].

Figura 14: Exemplo de *dashboard* em Power BI publicado na nuvem



Fonte: [19]

2.4 Excel

O Microsoft Office Excel é um software utilizado mundialmente tanto por empresas quanto por usuários comuns, principalmente para cálculos financeiros e estatísticos por conta da sua planilha eletrônica. Desenvolvido na década de 80 e com funcionamento nos dois principais sistemas operacionais existentes, Macintosh e Windows, permite desde a organização simples dos gastos pessoais ou de uma casa, até a gestão e controle de um complexo estoque de grandes empresas [20].

Entre outras funções com o Excel, estão organizar dados numéricos, textuais e gráficos em planilhas, resumindo os dados para uma melhor análise posterior. Através de suas funções pré-definidas conseguimos realizar cálculos estatísticos, financeiros, trigonométricos, lógicos, entre outros. O software possui a linguagem de programação Visual Basic que possibilita ao operador criar seus próprios gráficos, funções, tabelas e comandos avançados, aumentando a quantidade de atividades disponíveis pelo software [21].

2.5 Base de dados utilizada

A base de dados utilizada para realização desse trabalho foi fornecida por uma empresa privada do setor de seguros, com dados sobre as vendas da empresa para o período de janeiro de 2020 até abril de 2021. Para isso, foi realizada uma cópia do banco de dados SQL na *cloud*

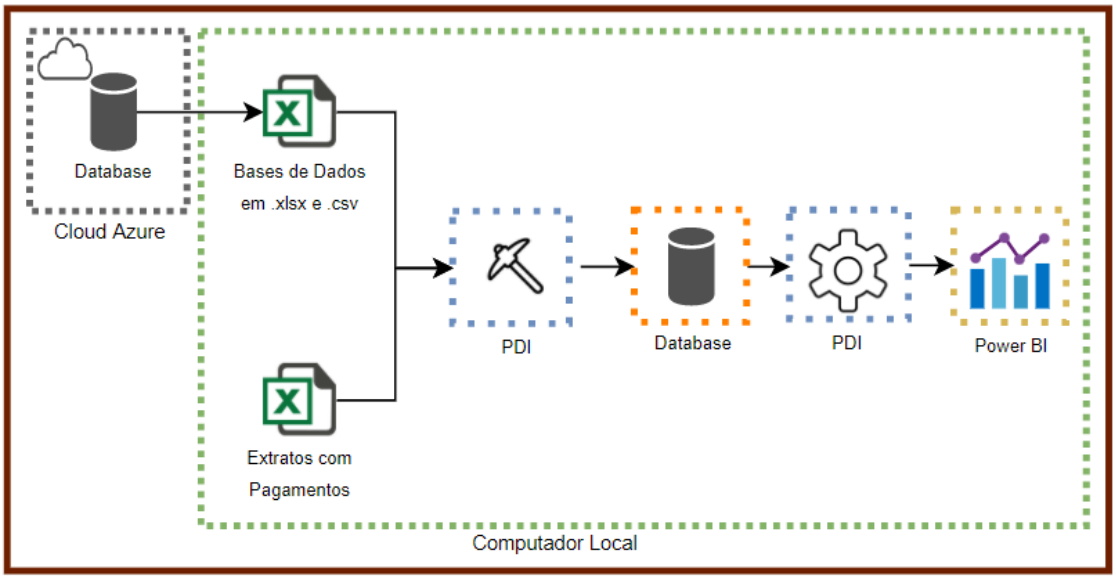
Azure da companhia para alguns arquivos em .csv e .xlsx para utilização do discente, visando o controle das informações que seriam passadas e também, evitar a conexão direta da ferramenta de ETL (PDI) utilizada neste trabalho aos servidores *cloud* da companhia. A "granularidade" dos dados está no nível de item vendido pela empresa, através de determinado fornecedor e para determinado cliente. Por conta da confidencialidade dos dados, algumas informações sensíveis foram suprimidas e outras, transformadas em anônimas (como o nome do cliente/fornecedor para "Cliente 1" e "Fornecedor 1") para impossibilitar o reconhecimento da carteira da empresa fornecedora dos dados.

3. DESENVOLVIMENTO

3.1 Arquitetura de dados

Para o desenvolvimento do trabalho a seguinte arquitetura foi utilizada, retratada pela Figura 15.

Figura 15: Arquitetura utilizada no trabalho



Fonte: Autoria própria

Pode-se observar através da Figura 15 a presença de duas estruturas principais: Cloud Azure e a máquina local do usuário.

A primeira, a Cloud Azure, é onde o banco de dados oficial da companhia está hospedado. Como já citado, por motivos de confidencialidade e segurança da informação, uma cópia das principais tabelas do banco de dados foi exportada para .csv e .xlsx pela área de TI da empresa e cedida ao discente. Já a segunda, o computador do usuário, é onde os arquivos .xlsx e .csv estão armazenados e onde a solução em PDI foi implementada e está operando, além doPower BI, ferramenta instalada e utilizada para análise e visualização dos dados.

Os dados cedidos pela empresa (arquivos .xlsx e .csv) foram inseridos em um banco de dados PostgreSQL para simular a conexão da ferramenta de ETL Pentaho ao banco de dados oficial da companhia. Para realizar a inserção, foi utilizado o próprio PDI, onde através de um conector nativo, extrai-se os dados dos arquivos e faz-se a transferência para o banco PostgreSQL.

Com os dados no banco já simulando a estrutura da empresa, o PDI foi conectado através de outro conector da ferramenta, onde informamos as credenciais de acesso como endereço,

porta utilizada, usuário e senha para autenticação. Assim, trazemos os dados que estão em tabelas do banco para dentro da ferramenta de ETL para início dos tratamentos, limpezas e processamento dos dados.

Por fim, com os processos realizados, criam-se as tabelas finais com os dados já cruzados entre si para análise final pelo usuário. Essas tabelas servirão de fonte para a ferramenta de visualização, no caso, o Power BI, para criação dos *dashboards* interativos.

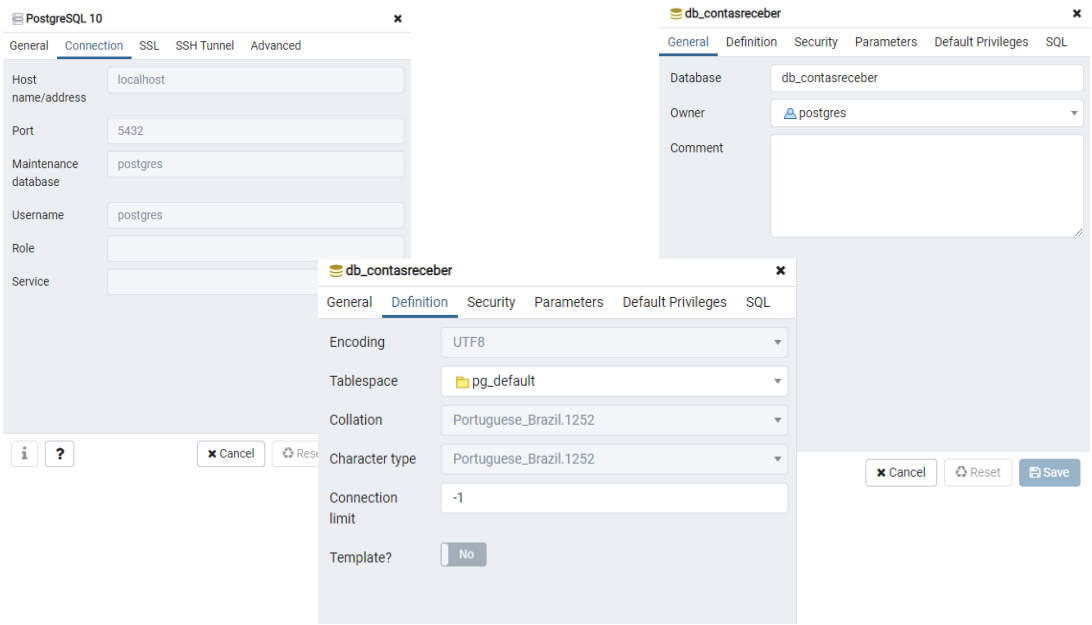
3.2 Criação do banco de dados PostgreSQL

A escolha pelo PostgreSQL para o banco de dados do trabalho foi motivada pela prévia utilização do discente durante o estágio, ou seja, já havia algum aprendizado que pudesse ser reaproveitado a fim de garantir melhores resultados com o desenvolvimento da ferramenta.

Além disso, como o projeto não poderia se conectar diretamente ao banco de dados oficial da empresa, havia a necessidade da criação de um BD intermediário para simular a estrutura oficial da companhia.

Assim, criou-se uma instância simples de banco de dados na ferramenta do PostgreSQL com codificação padrão (UTF-8) para receber as tabelas exportadas pela equipe de TI do banco original no formato de Excel. A Figura 16 apresenta três telas do banco com as suas configurações.

Figura 16: Configurações do banco de dados.

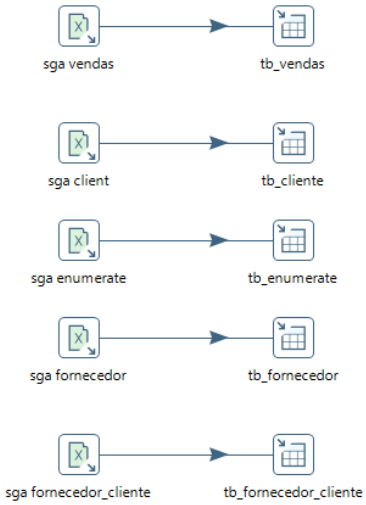


Fonte: Autoria própria

3.3 Desenvolvimento do ETL

De posse dos arquivos fornecidos pela equipe de TI com visões atuais do banco de dados original, utilizou-se o Pentaho Data Integration para inserir os dados desses arquivos para o banco de dados PostgreSQL criado pelo discente, através do fluxo criado e representado na Figura 17.

Figura 17: Fluxo que migra tabelas do banco oficial da companhia (Excel) para o banco de dados do TG



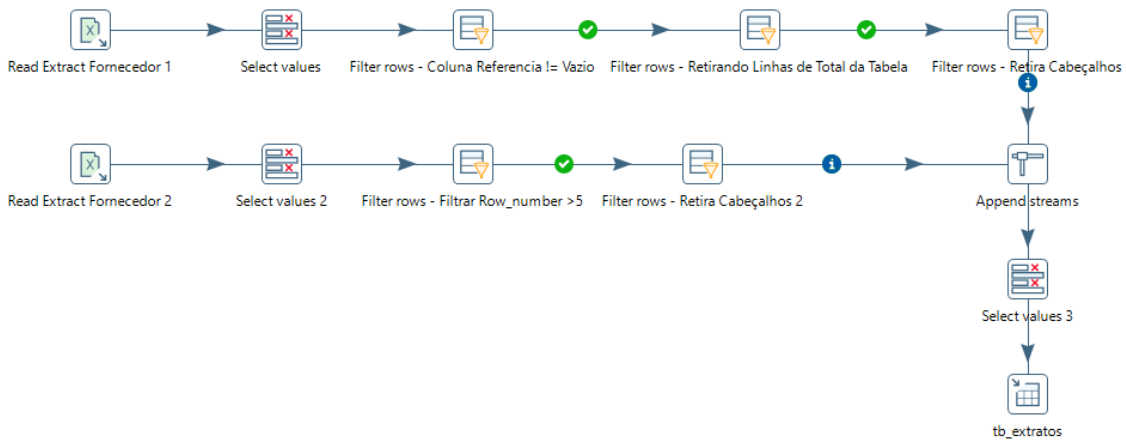
Fonte: Autoria própria

O fluxo na figura 17 é responsável por pegar cada um dos arquivos em Excel e inserir os dados em diferentes tabelas do banco. Essas tabelas são a base do banco de dados do trabalho e envolvem dados de vendas, dos clientes e dos fornecedores.

Após a obtenção das tabelas de vendas e dos dados cadastrais de clientes e fornecedores, obteve-se as tabelas de extrato, que comprovam o recebimento dos valores dos itens vendidos. Para isso, um novo fluxo foi criado (Figura 18), onde todos os arquivos de extrato dos fornecedores que estão na pasta do repositório são lidos e os dados imputados em uma tabela final de extratos. A Figura 18 apresenta o fluxo de obtenção dos dados dos extratos.

É importante destacar que é nesse momento que a automação de leitura dos extratos é realizada, visto que qualquer arquivo em Excel que seja colocado na pasta e tenha mesmo *layout* do fornecedor, será lido e enviado para o banco.

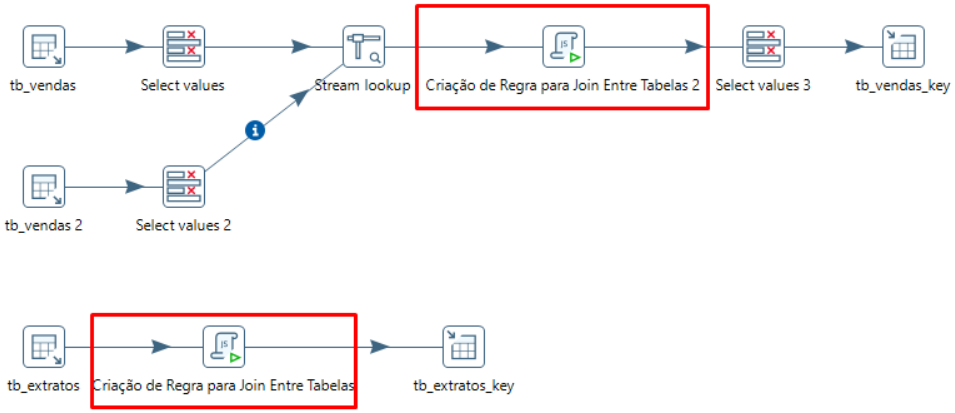
Figura 18: Fluxo que migra os extratos (Excel) para o banco de dados do TG



Fonte: Autoria própria

Com os dados de vendas e dos extratos no banco de dados, é criada uma chave para comparação desses dois itens. A chave consiste na utilização de dois códigos presentes na venda, antecedidos por uma letra. Essa letra (“A” ou “E”) é utilizada quando um dos códigos está ausente (“A”) ou, caso os dois códigos estejam presentes, (“E”). Não há casos em que nenhum dos códigos estejam presentes. O fluxo responsável por criar essa chave tanto na base de vendas quando na base de extratos é apresentado na Figura 19. As atividades responsáveis pela criação dessa regra de comparação estão identificadas com um retângulo.

Figura 19: Fluxo que cria as chaves de comparação para os itens vendidos e extratos pagos

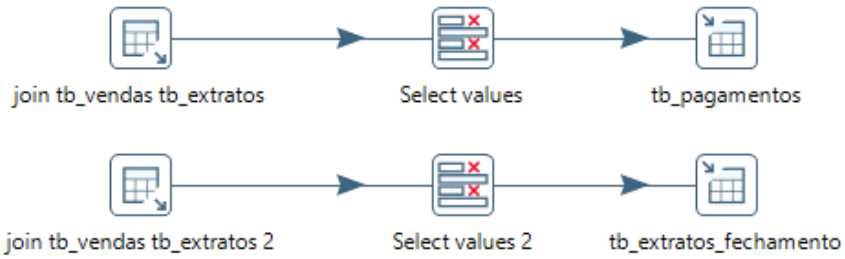


Fonte: Autoria própria

Por fim, as informações da tabela de vendas e de extratos são cruzadas para criação da tabela de pagamentos e tabela de conferência dos extratos, representadas na Figura 20. São feitos

ambos os cruzamentos, tanto para conhecimento dos itens que foram vendidos e já foram recebidos pelo financeiro, quanto para conferência inversa dos extratos (caso haja alguma linha de pagamento no extrato que não está no controle dos itens vendidos).

Figura 20: Fluxo que cria a tabela de pagamentos após o cruzamento das tabelas de vendas e extratos

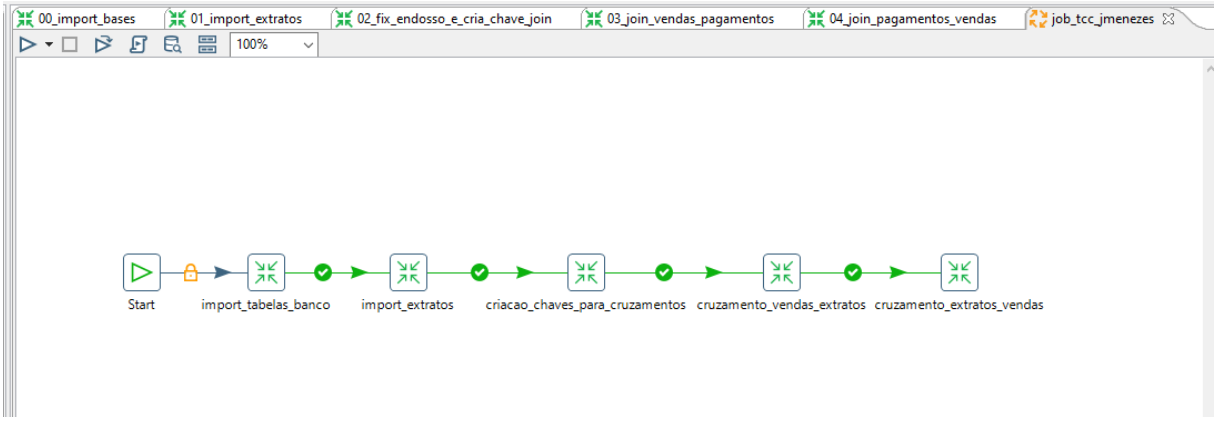


Fonte: Autoria própria

Com os cruzamentos feitos, temos as tabelas necessárias para realização das análises e criação visuais através do Power BI.

Por fim, foi criado um *job* que executa todas as transformações criadas no Pentaho de forma sequencial (Figura 21).

Figura 21: *Job* criado com execução de todas as transformações feitas no Pentaho



Fonte: Autoria própria

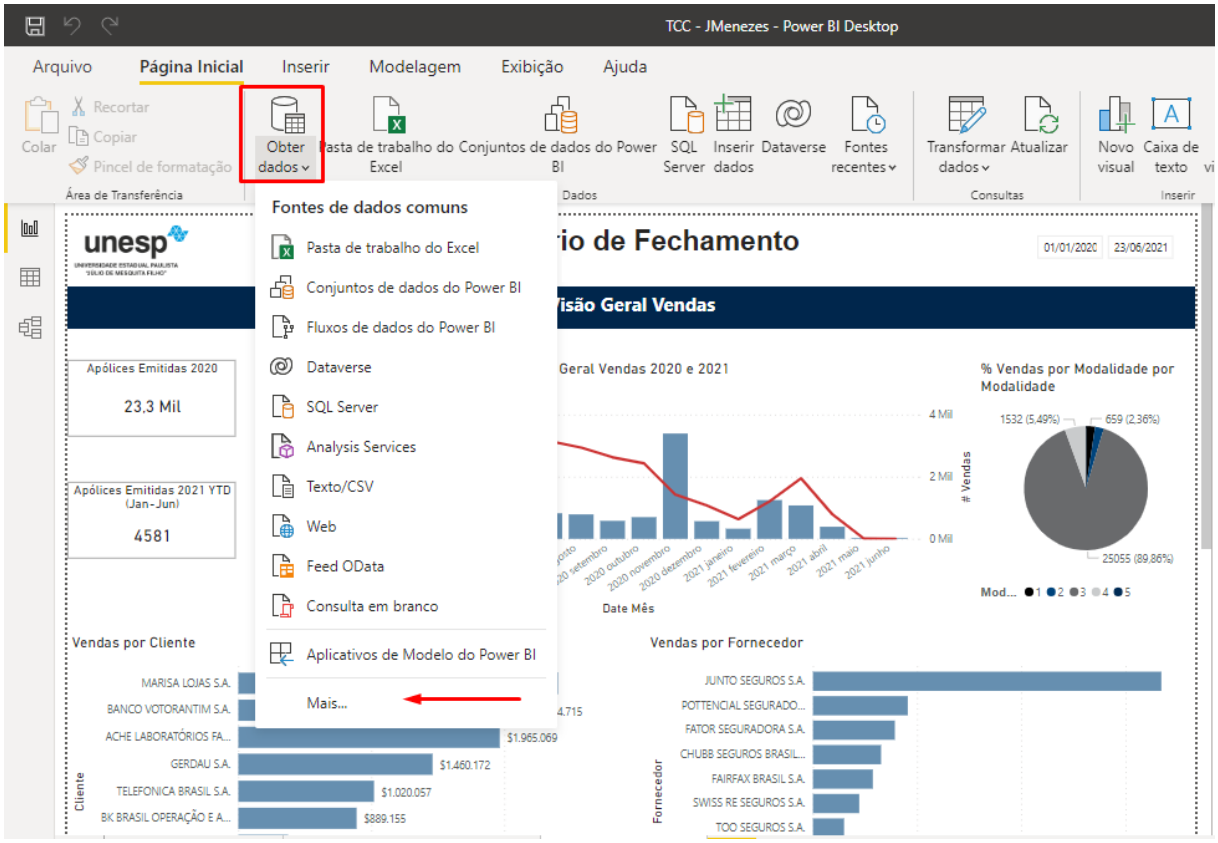
Para a execução facilitada do *job*, foi criado um arquivo *.bat* com o diretório do arquivo do Job. Basicamente, o arquivo chama o Kitchen do Pentaho (previamente explicado no trabalho) para a execução do *job*, que consequentemente executa as transformações desenvolvidas e implementadas pelo discente.

3.4 Criação dos *Dashboards* em Power BI

Após a criação das tabelas no banco de dados, utilizou-se o Power BI para criação do relatório analítico, com gráficos e tabelas que apresentam as informações necessárias para conferência das vendas.

No Power BI, na aba de "Página Inicial" e no ícone de "Obter dados", é possível selecionar qual será a conexão utilizada para aquisição das informações pela ferramenta, como mostra a Figura 22. Ao clicar em "Mais...", outras fontes de dados são apresentadas, além das mais comuns.

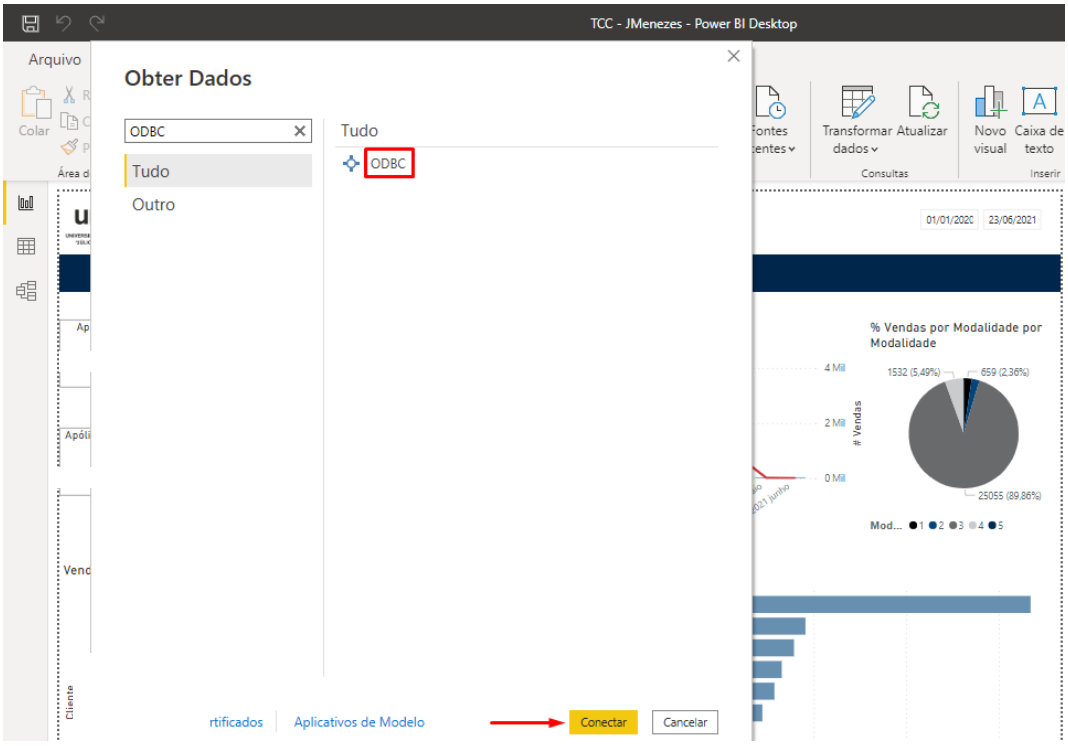
Figura 22: Conexão com tabelas do banco de dados na ferramenta do Power BI



Fonte: Autoria própria

Após selecionar o botão "Mais..." para seleção de outras fontes, deve-se escolher o item "ODBC" – *Open Database Connectivity* - para realização da conexão com o banco de dados local criado (Figura 23). O termo ODBC é utilizado para se referir a interface de conexão da Microsoft que permite que aplicações (como o Power BI) se conecte a bancos de dados através de um método padrão.

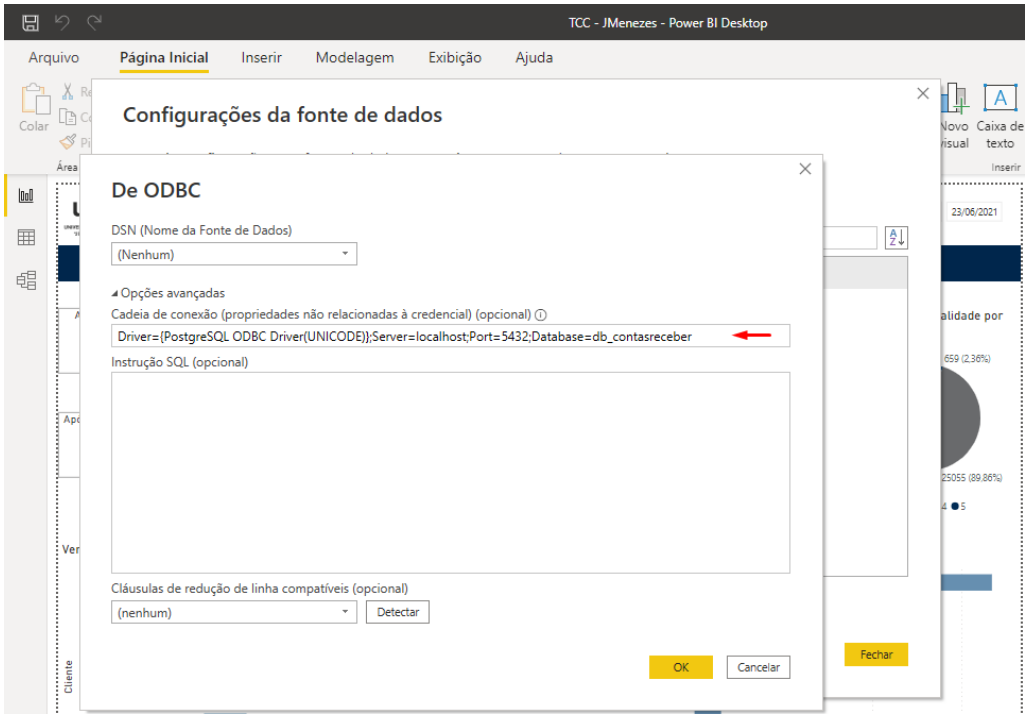
Figura 23: Seleção do conector correto para o banco de dados local criado



Fonte: Autoria própria

Por se tratar de um banco de dados local, o campo “DSN” é deixado como “(Nenhum)” e é necessária a escrita da chave de conexão com o banco no campo “Cadeia de conexão” da ferramenta. A Figura 24 apresenta a imagem com os campos selecionados e a cadeia de conexão descrita para a conexão no banco de dados utilizado pelo discente.

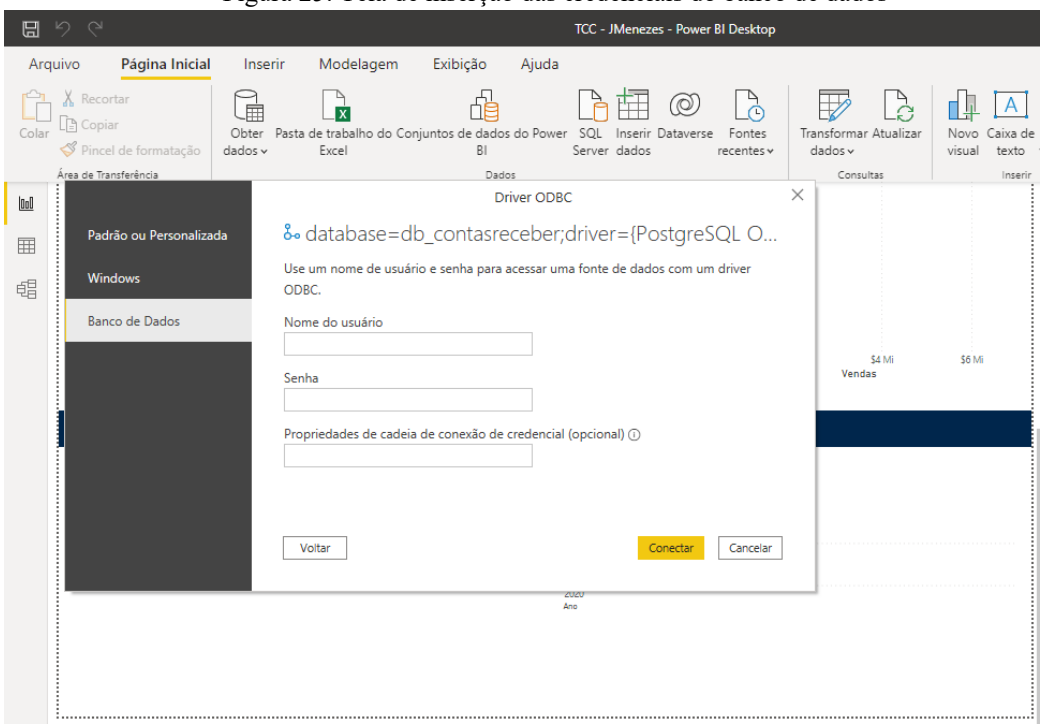
Figura 24: Conexão do Power BI com o banco local criado



Fonte: Autoria própria

Após a realização da conexão, os dados de usuário e senha para o banco de dados são inseridos na ferramenta nos campos da ferramenta (Figura 25).

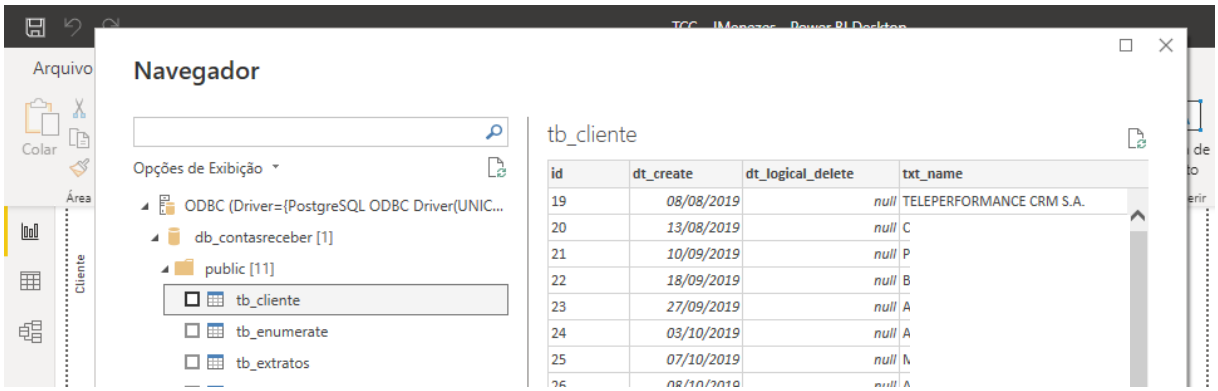
Figura 25: Tela de inserção das credenciais do banco de dados



Fonte: Autoria própria

Com a inserção correta das credenciais do banco, a ferramenta disponibiliza todas as tabelas que foram encontradas nesse banco de dados, cabendo ao usuário a seleção das que deseja utilizar no navegador da ferramenta. A Figura 26 apresenta a visualização desse navegador.

Figura 26: Tela de seleção das tabelas do banco de dados



Fonte: Autoria própria

As tabelas selecionadas e seus conteúdos estão apresentadas na Tabela 1.

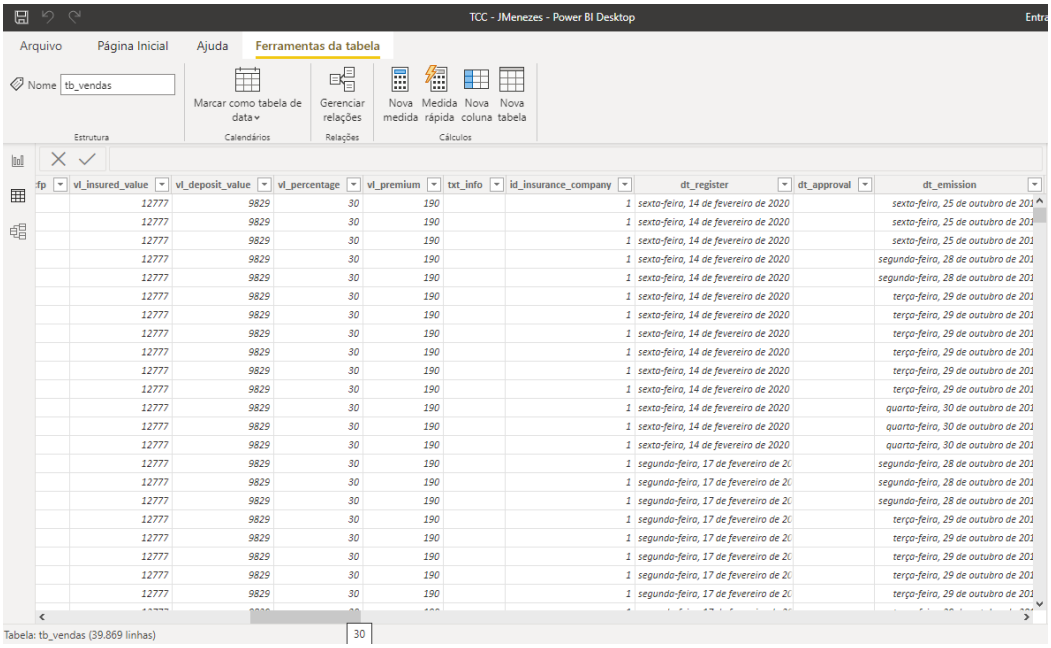
Tabela 1: Tabelas utilizadas na criação do *Dashboard*

Nome da Tabela	**Descrição**
Data	Tabela criada dentro do próprio Power BI para servir como chave no relacionamento de datas entre as vendas e os recebimentos
TB_Vendas	Tabela do banco de dados que contém todos os itens vendidos pela companhia
TB_Clientes	Tabela do banco de dados que contém dados cadastrais dos clientes
TB_Extratos	Tabela do banco de dados que contém dados dos pagamentos realizados
TB_Pagamentos	Tabela do banco de dados que contém todos os pagamentos recebidos por meio do cruzamento entre as vendas e os extratos da companhia
TB_Fornecedor	Tabela do banco de dados que contém dados cadastrais dos fornecedores da companhia

Fonte: Autoria própria

Com as tabelas “baixadas” no Power BI (Figura 27), iniciou-se a criação dos indicadores e gráficos para utilização pela empresa.

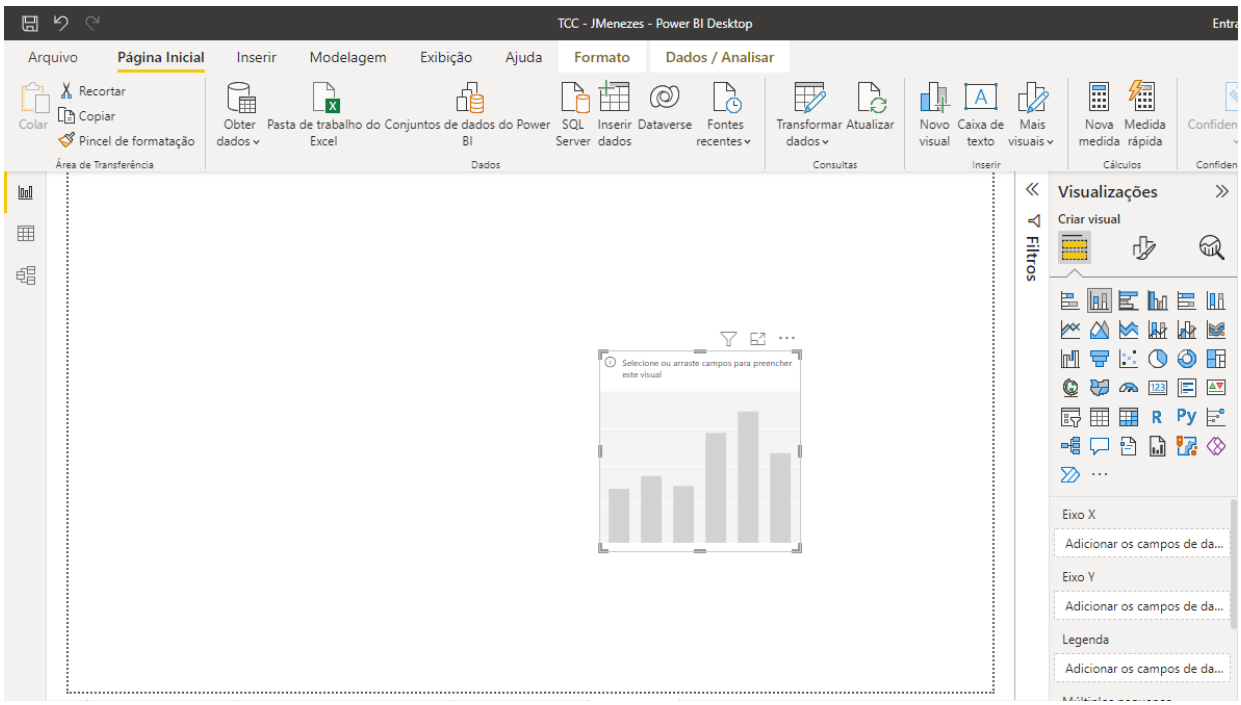
Figura 27: Recorte da tabela TB_Vendas já incorporada no Power BI



Fonte: Autoria própria

Para a criação dos indicadores e gráficos, seleciona-se a visualização desejada no painel do Power BI e ela automaticamente é inserida na página da ferramenta. A Figura 28 apresenta como exemplo, a seleção da visualização “Gráfico de colunas empilhadas” do Power BI.

Figura 28: Criação de um indicador no Power BI

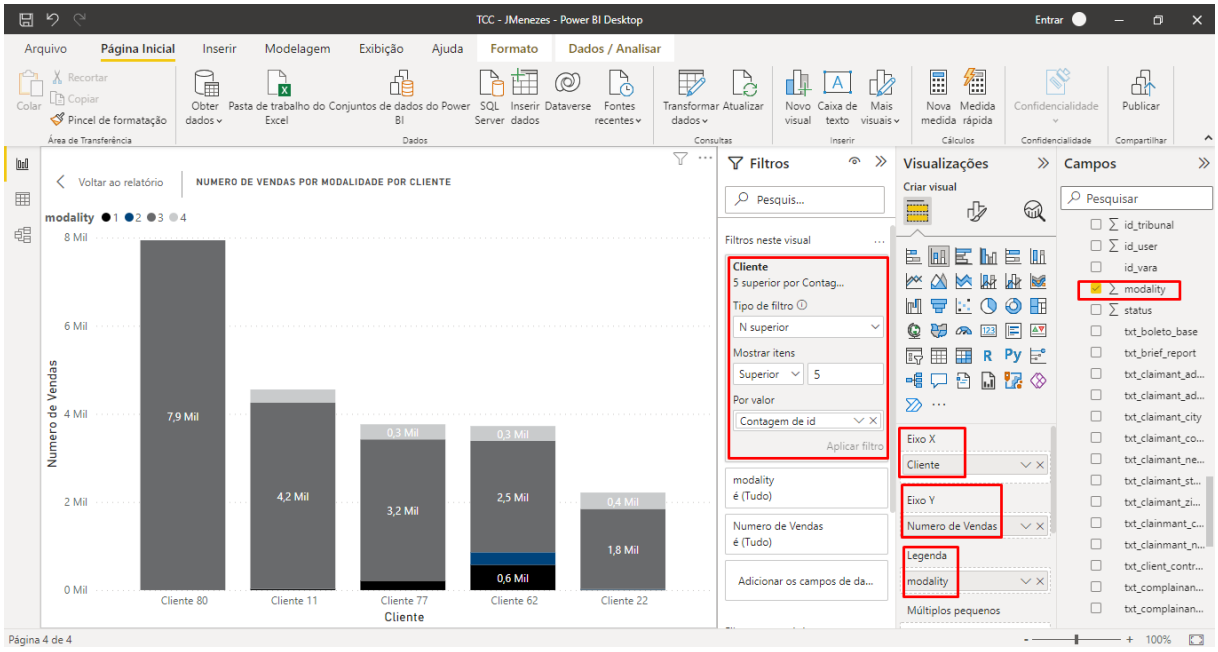


Fonte: Autoria própria

Ao criar o indicador, é necessário selecionar as colunas da tabela que se deseja analisar e colocar nos parâmetros. Como ilustração, vamos analisar o número de itens vendidos por modalidade, por cliente. Assim, selecionou-se a coluna que informa o cliente da tabela de clientes e a posicionou no eixo X. A coluna com o número da venda da tabela de vendas e a posicionou no eixo Y, e por fim, a modalidade da venda também da tabela de vendas, como legenda. Para selecionar, é só arrastar a coluna da paleta de tabelas no canto direito da ferramenta e soltar na paleta de visualização, com o indicador selecionado.

Além disso, para melhor visualização, foi filtrado apenas os 5 maiores clientes que compraram. O resultado é apresentado na Figura 29. O número de clientes filtrados na paleta de “Filtros” e os campos das tabelas selecionados na paleta de “Visualizações”, além das colunas da tabela na paleta de “Campos” foram destacados com um retângulo na imagem.

Figura 29: Indicador genérico criado para ilustração de como criar gráficos no Power BI



Fonte: Autoria própria

4. RESULTADOS e DISCUSSÕES

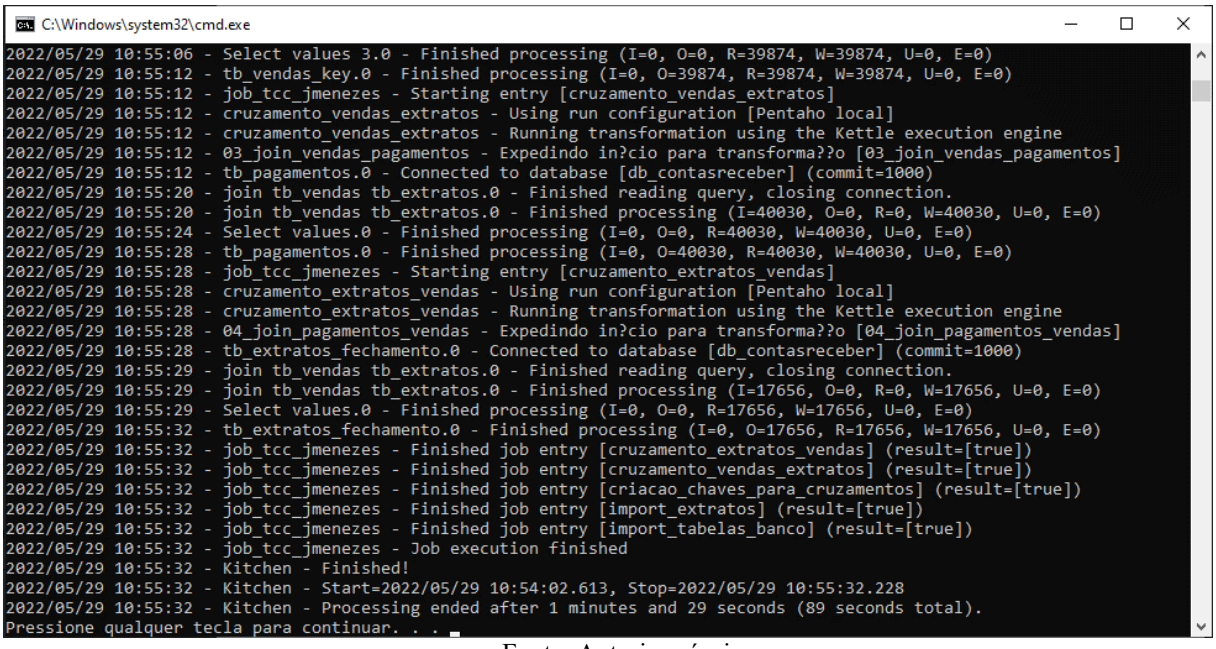
4.1 ETL via Pentaho

Para execução do ETL criado dentro do software Pentaho Data Integration, foi desenvolvido um arquivo *.bat* que executa os arquivos do programa de forma externa ao software, ou seja, sem ser necessário sua abertura. Assim, para o usuário fazer todo o ETL implementado funcionar, ele apenas precisa executar esse arquivo criado, que contém o seguinte código:

```
CALL C:\PDI 9.0\Kitchen.bat/file :C:\Users\ JMenezes \Documents\T G  - JMENEZES\
\ktrs\job_tcc_jmenezes.kjb
PAUSE
```

Antes do desenvolvimento do ETL, o funcionário da empresa levava em torno de 15h por semana para poder comparar todas as vendas da empresa com os extratos recebidos, a fim de validar os itens que já foram pagos pelos clientes e gerar os indicadores para a alta liderança. O processo dependia muito tempo, pois não havia agilidade na utilização do Excel, muitas vezes utilizado apenas como uma planilha manual (sem fórmulas e sem macros). Com a elaboração do ETL, o tempo de comparação das informações passa a ser de 90 segundos, conforme apresenta a Figura 30.

Figura 30: Print de execução do arquivo .bat que executa o job com todas as transformações criadas



Fonte: Autoria própria

A Figura 30 é o recorte final do *log* de execução do arquivo *.bat*, onde podemos verificar na penúltima linha o tempo de execução dessa rodada (89 segundos) e um pouco mais acima, que a execução das transformações foram bem-sucedidas. A execução desse arquivo *.bat* é necessária toda vez que novos extratos forem recebidos pelo analista. Eles devem ser posicionados na pasta correspondente do seu fornecedor para que o *layout* esteja condizente com o que o Pentaho foi programado.

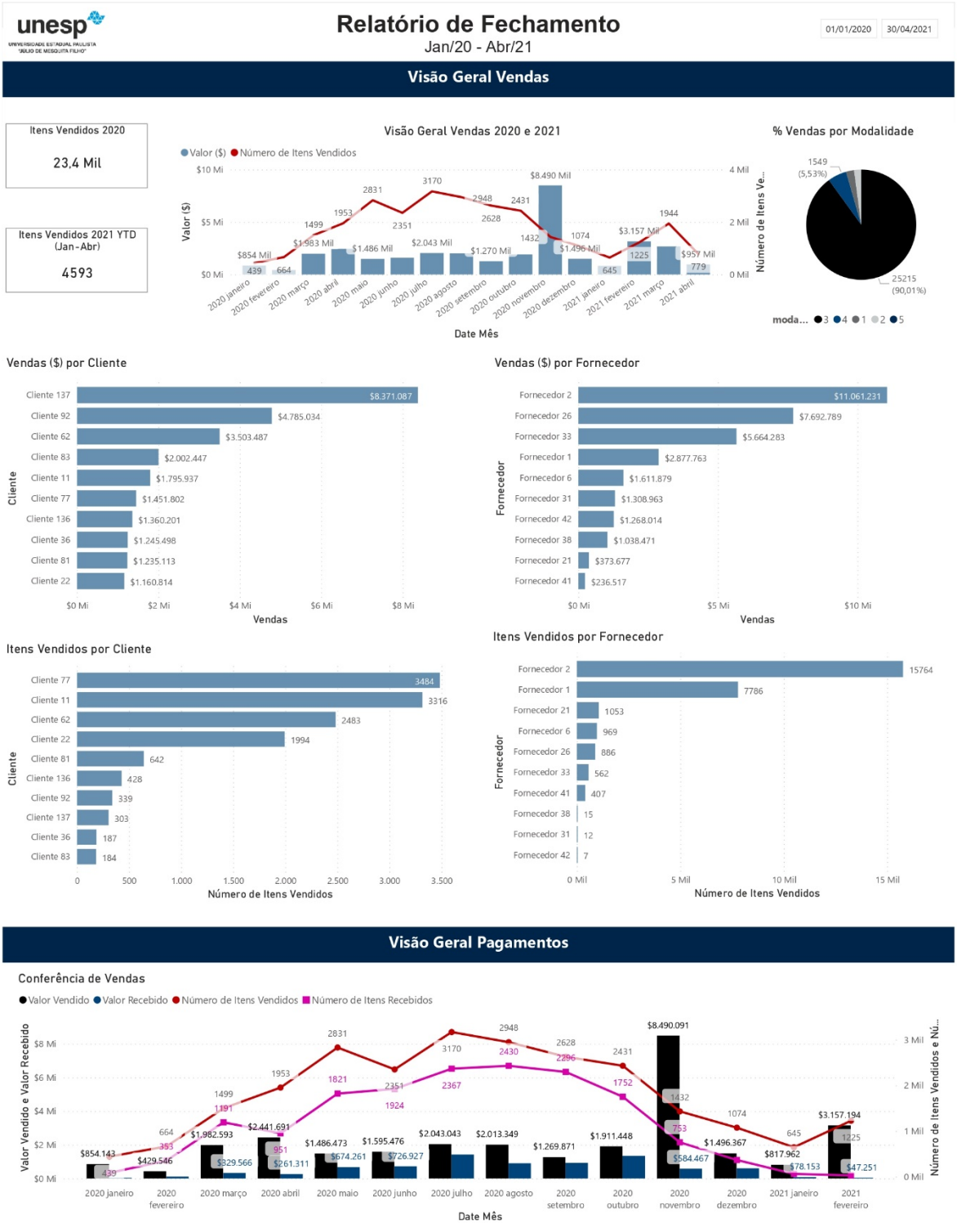
Ao final da execução, as tabelas do banco de dados foram atualizadas com as informações mais recentes e já estão disponíveis para consulta e análise do analista via Power BI. Anteriormente ao desenvolvimento da solução, o analista gastava em torno de 70h mensais entre pegar as informações das vendas com a TI, pegar as informações de itens recebidos com os fornecedores, analisar as informações utilizando excel (comparando linha a linha) e gerar as análises necessárias para reportar aos superiores.

4.2 Análise dos dados via Power BI

Com as tabelas integradas ao Power BI, desenvolveu-se o que chamamos de "*One Page Report*" em *business intelligence*: um relatório de uma página que normalmente se estende em altura, dando todas as informações gerenciais necessárias para a análise e conclusões sobre o problema. A Figura 31 apresenta a visão macro desse relatório.

O Relatório de Fechamento apresentado na Figura 31 contém duas seções principais: a primeira, Visão Geral Vendas, que apresenta ao usuário informações sobre número de itens vendidos pela companhia e o valor de receita que foi gerado por eles. Além disso, também possui indicadores com os principais clientes e fornecedores; já a segunda seção, apresenta a visão das vendas em comparação com os recebimentos, assim o analista e a alta liderança conseguem saber o que foi vendido e recebido pela companhia. Todo o painel é interativo, o que significa que ao clicar em um cliente ou em um fornecedor específico, o restante dos gráficos e indicadores são filtrados para apresentar os dados desse cliente/fornecedor escolhido.

Figura 31: Visão macro do *One Page Report* criado no Power BI



Fonte: Autoria própria

A Tabela 2 apresenta então, todos os indicadores criados e suas descrições.

Tabela 2: Tabela com todos indicadores e gráficos criados

Nome do Indicador	**Descrição**
Itens Vendidos 2020	Indicador do tipo cartão, que contém o total de itens vendidos apenas no ano de 2020
Itens Vendidos 2021 YTD	Indicador do tipo cartão, contendo os itens vendidos em 2021 YTD – *Year to Date*, ou seja, do começo do ano até o momento do recorte dos dados, feito em abril de 2021
Visão Geral das Vendas 2020 e 2021	Indicador do tipo gráfico de colunas empilhadas com linha que apresenta a visão geral das vendas por mês e/ou por ano desde o começo de 2020 até abril de 2021
% Vendas por Modalidade	Indicador do tipo gráfico de pizza que apresenta a porcentagens de itens vendidos dentro das 4 modalidades da empresa
Vendas ($) por Cliente	Indicador do tipo gráfico de barras clusterizado que contém o valor em moeda de venda por cliente*.
Vendas ($) por Fornecedor	Indicador do tipo gráfico de barras clusterizado que contém o valor em moeda de venda por fornecedor*.
Itens Vendidos por Cliente	Indicador do tipo gráfico de barras clusterizado que contém o número de itens vendidos por cliente*.
Itens Vendidos por Fornecedor	Indicador do tipo gráfico de barras clusterizado que contém o número de itens vendidos por fornecedor*.
Valor Vendido e Valor Recebido por Ano e Mês	Indicador do tipo gráfico de colunas clusterizado que contém por mês os valores vendidos e que foram recebidos (na conferência com os extratos).

Fonte: Autoria própria

*Os clientes e fornecedores estão descaracterizados por serem dados confidenciais.

Sempre que novos extratos são recebidos pelo analista, após a execução do arquivo *.bat*, o mesmo deve atualizar o Power BI para sincronizar com as informações mais recentes que estão disponíveis no banco de dados. A Figura 32 destaca o botão que atualiza a ferramenta.

Figura 32: Botão que atualiza o Power BI com os dados mais recentes nas tabelas do banco de dados

Fonte: Autoria própria

Com as tabelas atualizadas, os indicadores automaticamente apresentam os dados mais recentes. Cada um dos indicadores criados para análise serão agora explicados e retratados a seguir.

4.2.1 Visão Geral Vendas

A primeira seção do *dashboard* apresenta indicadores relacionados ao número de itens vendidos e quanto esses geraram de receita para a empresa. A Figura 33 apresenta os dois indicadores do tipo cartão.

Figura 33: Indicadores do tipo cartão com os números de itens vendidos

Fonte: Autoria própria

Os indicadores do tipo cartão são utilizados para ressaltar números importantes, sem que o usuário necessite desprender tempo para analisar outros indicadores para chegar em conclusões rápidas.

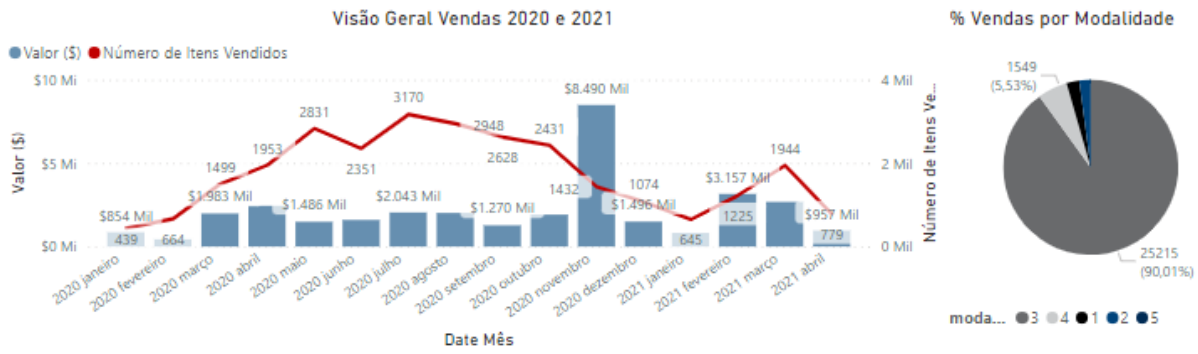
O primeiro cartão apresenta os itens vendidos em 2020, enquanto o segundo cartão os itens vendidos em 2021 até o momento de amostragem dos dados. O termo YTD é um termo utilizado frequentemente no mundo corporativo e que se refere a "*Year to Date*", ou seja, do começo do ano até o atual momento. Como os dados fornecidos ao discente para o trabalho de

graduação foram até abril/21, o termo é utilizado mas a data de referência é explicitada na legenda do cartão (Jan-Abr).

Analisando os dados, reparamos que em 2020 a média de itens vendidos por mês é de aproximadamente 1.942 itens. Consequentemente, não pensando em aumento de receita para o próximo ano, era esperado por volta de 7700 contra apenas 4.573 itens de fato vendidos (queda de mais de 40% no esperado).

A Figura 34 apresenta o gráfico de colunas com linha e gráfico de pizza, que trazem uma visão mês a mês de venda desde o início de 2020 e da quebra entre as modalidades de venda, respectivamente.

Figura 34: Gráfico de colunas com linha e de pizza, trazendo visão mensal das vendas e da quebra em modalidades



Fonte: Autoria própria

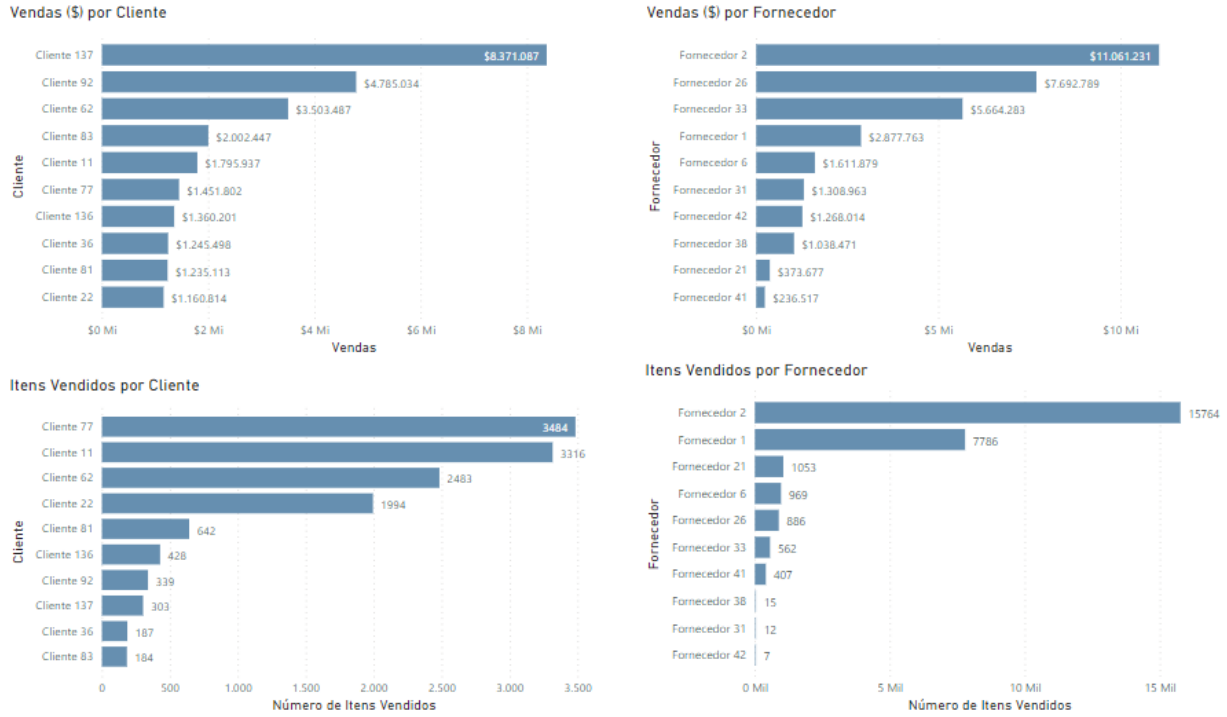
No gráfico com colunas e linha, as colunas apresentam o valor em moeda e a linha, o número de itens vendidos no respectivo mês. Os valores que estão sem unidade (“Mil”) são referentes ao número de itens vendidos e os que estão com a unidade dos milhares, o valor monetário. Pode-se perceber que o mês de novembro de 2020 é um mês atípico, onde embora o número de itens vendidos seja menor que dos meses anteriores o valor monetário supera e muito, devido a modalidade das vendas. O gráfico de pizza traz que dos itens vendidos, a grande maioria se concentra na modalidade 3, com quase 90% de representatividade.

Analisando agora a Figura 35, ela traz os gráficos com barras clusterizado, que apresentam os itens vendidos e o valor vendido com quebras por cliente e fornecedor.

Nota-se que embora o valor ($) e os itens vendidos por cliente sejam mais pulverizados, pelo lado dos fornecedores, o Fornecedor 1 e o Fornecedor 2 concentram 84% do número de itens vendidos e 56% do valor vendido.

Já para os clientes, vale destacar a representatividade dos Clientes 137, 92, 62 e 83 com relação ao valor vendido com 31%, 18%, 13% e 7%, respectivamente, e dos Clientes 77, 11, 62 e 22 com relação ao número de itens vendidos com, respectivamente, 26%, 25%, 19% e 15%.

Figura 35: Gráficos de barras com quebras de valor e número de itens vendidos para clientes e fornecedores



Fonte: Autoria própria

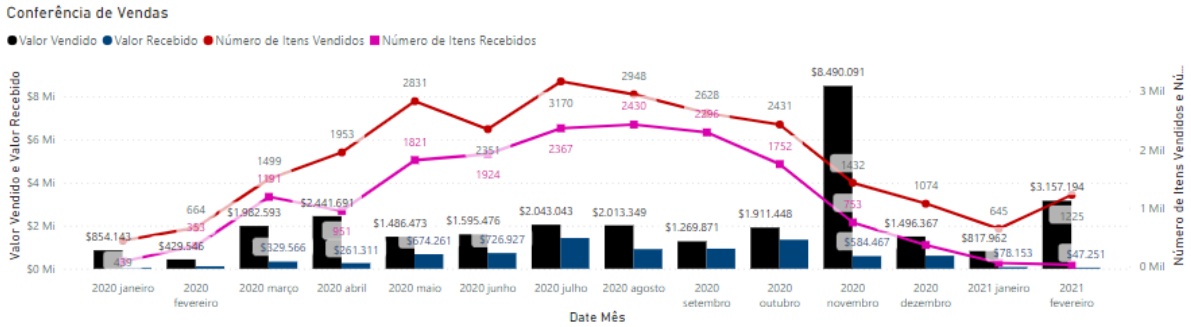
4.2.2 Visão Geral Pagamentos

Para a conferência dos itens, foi plotado um gráfico de colunas e linhas, onde as colunas representam o valor vendido e recebido, e as linhas, o número de itens vendidos e recebidos. A Figura 36 apresenta a visualização criada para esses dados.

No gráfico, os dados estão separados por mês, onde as colunas da esquerda representam o valor vendido e as da direita, o valor recebido. Para as linhas, o marcador circular representa o número dos itens vendidos e o marcador em forma de quadrado, o número dos itens que foram recebidos para aquele mês.

Pode-se concluir que a empresa precisa reforçar seus processos de cobrança, visto que em nenhum dos meses apresentados o valor recebido está próximo do valor vendido.

Figura 36: Gráfico de colunas e linhas utilizados para conferência dos valores vendidos e recebidos.



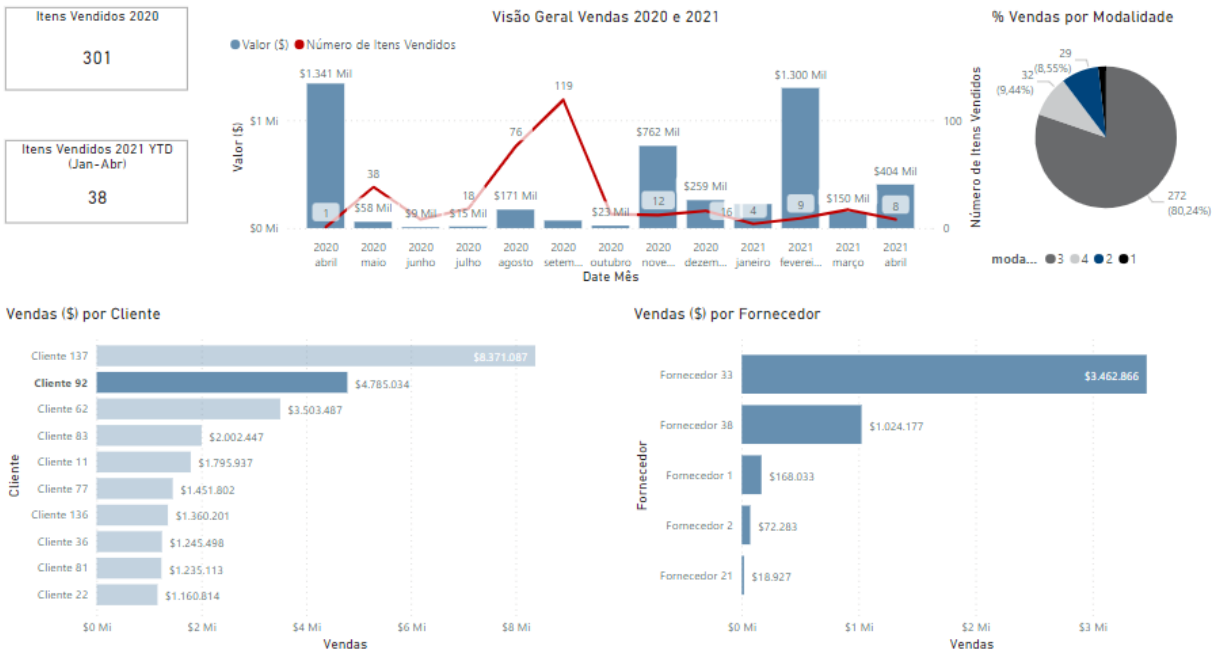
Fonte: Autoria própria

4.2.3 Interatividade do Power BI

A ausência de filtros no *dashboard* é dada pela alta interatividade da ferramenta Power BI, visto que qualquer um dos gráficos apresentados está com relacionamento interno com os outros. Dessa forma, caso deseje analisar com mais especificidade algum cliente, fornecedor ou mês, é necessário apenas selecionar esse item desejado que o restante do painel será filtrado de acordo com a seleção.

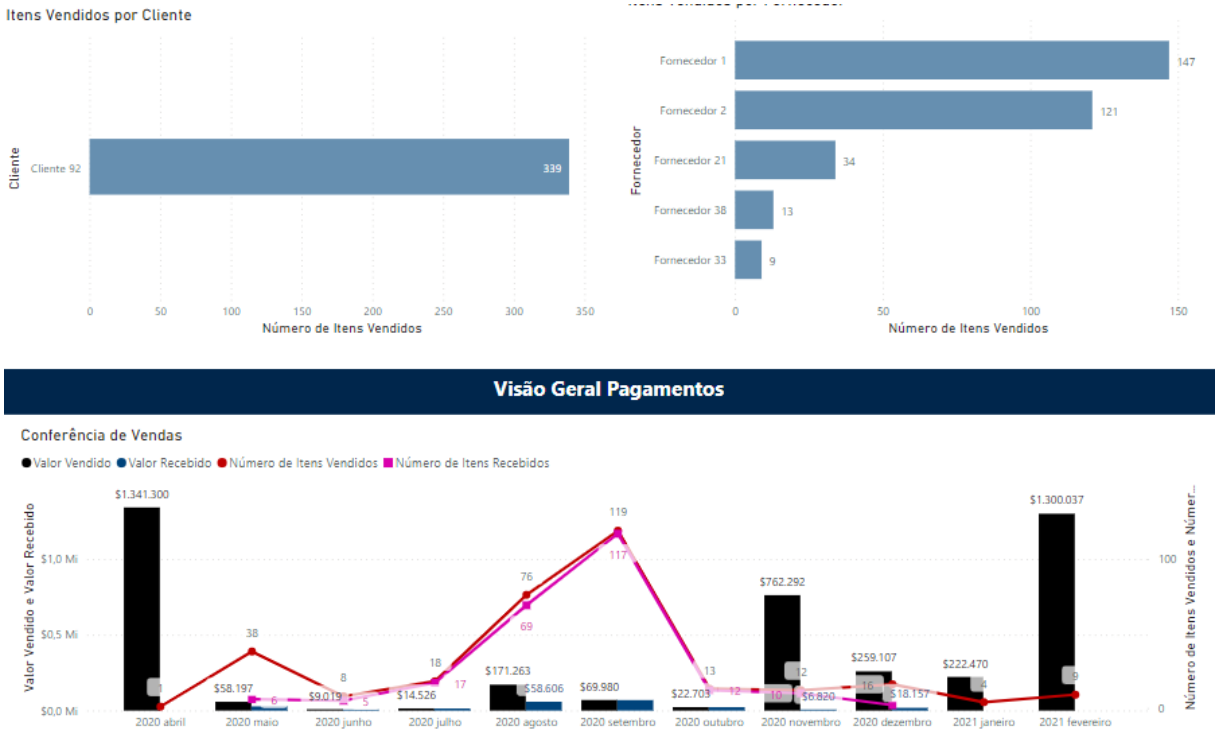
As Figuras 37 e 38 mostram exatamente essa visão, apresentando o filtro aplicado para o Cliente 92.

Figura 37: Primeira metade do *dashboard*, com filtro aplicado ao selecionar o Cliente 92 no gráfico de colunas horizontais.



Fonte: Autoria própria

Figura 38: Segunda metade do *dashboard*, com filtro aplicado ao selecionar o Cliente 92 no gráfico de colunas horizontais



Fonte: Autoria própria

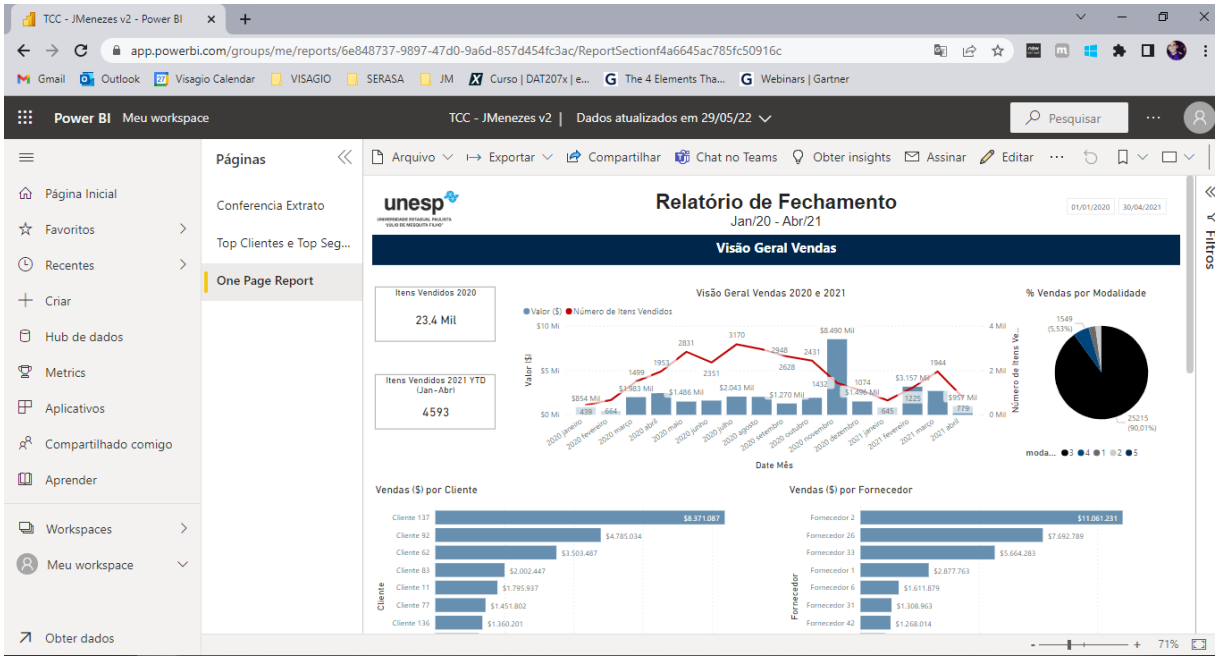
Note que todos os outros indicadores se moldam à seleção que foi realizada, apresentando os dados apenas para o Cliente 92 selecionado.

4.2.4 Power BI Service – Acesso *Online*

O Power BI permite a publicação do seu *dashboard* na plataforma *online*, onde dentro de uma organização é possível ter acesso às informações sem necessariamente ter o software instalado em sua máquina local. A Figura 39 apresenta a visualização do *site*.

Dessa forma, outros funcionários da empresa e a liderança, conseguem analisar as informações e se atualizar de forma muito mais rápida e ágil, logo quando o analista financeiro publicar as informações.

Figura 39: Visualização da plataforma *online* do Power BI, onde o *dashboard* foi publicado



Fonte: Autoria própria

Para publicar o *dashboard* é necessário selecionar o botão “Publicar”, como mostra a Figura 40. Após a seleção, é necessário informar em qual local da plataforma deseja-se que os dados sejam inseridos.

Figura 40: Botão de publicação do *dashboard*

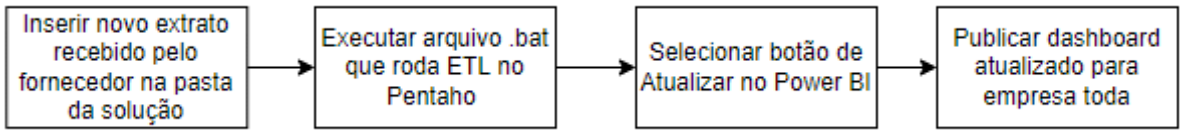
Fonte: Autoria própria

4.3 Fluxograma Operacional da Solução Implementada

A Figura 41 apresenta o fluxograma do processo que o analista deverá realizar para utilizar a ferramenta implementada. Esse fluxograma descreve as ações necessárias para o analista rodar toda a solução desenvolvida e ter em mãos os dados mais atualizados de vendas

e extratos disponíveis em seu Power BI e também, no Power BI Service (ferramenta *online* que disponibiliza o *dashboard* via nuvem para toda a companhia).

Figura 41: Processo a ser desempenhado pelo analista para atualização dos indicadores



Fonte: Autoria própria

5. CONCLUSÃO

O desenvolvimento desse trabalho mostrou que os softwares de *business intelligence* levam agilidade não só para o analista mas para toda a empresa, visto que antes os relatórios que eram elaborados e direcionados apenas para algumas pessoas via *e-mail*, agora podem ser acessados por qualquer um dentro da empresa com as licenças do Power BI através da internet – apenas acessando o relatório via *browser*.

Os resultados mostraram quão facilitado se tornou o dia a dia do analista que previamente utilizava o Excel para conferir manualmente linha por linha os itens vendidos, e agora, possui uma solução que realiza a comparação automática. O processo completo que levava aproximadamente 70h por mês, passa a ser realizado em torno de 5h a 10h por mês, possibilitando ao analista empregar mais tempo agora analisando os dados para solucionar outros problemas dentro da empresa. O ganho de tempo pode ser utilizado para outras tarefas ou então, para gerar hipóteses de como alavancar os resultados da empresa.

Além disso, indicadores diferentes que possam surgir na empresa, poderão facilmente ser implantados em outras páginas do *dashboard* e gerar insumos importantes para a mesma.

Observa-se também, a completude da solução desenvolvida com PostgreSQL, Pentaho Data Integration e Power BI, que juntas são capazes de desenvolver uma arquitetura simples, porém poderosa, e capaz de resolver problemas complexos de tratamento de dados dentro das empresas.

Pode-se concluir ainda, que a automação de processos tem se mostrado extremamente valiosa e está cada vez mais presente no cotidiano das companhias, visto sua capacidade de eliminar processos repetitivos e trazer mais qualidade para o trabalho desempenhado.

REFERÊNCIAS

GEYER-KLINGEBER, J.; NAKLADAL, J.; BALDAUF, F.; VEIT, F. Process Mining and Robotic Process Automation: A Perfect Match. Paper presented at 16th International Conference on Business Process Management 2018, Industry Track Session, Sydney, Australia.

XAVIER, C.; MOREIRA, F. Agile ETL. Elsevier, Procedia technology, 2013, Vol.9, pp.381-387.

MICROSOFT POWER BI. Power BI. Disponível em https://powerbi.microsoft.com/pt-br/ Acessado em 07 de outubro de 2020.

ORACLE BRASIL. Banco de Dados. Disponível em https://www.oracle.com/br/database/what-is-database.htm. Acessado em 26 de setembro de 2020.

ELMASRI, Ramez; NAVATHE, Shamkant B. SISTEMAS DE BANCO DE DADOS. 6. ed. São Paulo: Addison Wesley, 2011.

WADE, B.; CHAMBERLIN, D.; IBM Relational Database Systems: The Early Years. IEEE Annals of the History of Computing, v. 34, n. 4, p. 38-48, 2012.

GRAD, B.; BERGIN, T. J. History of Database Management Systems. IEEE Ann. Hist. Comput., vol. 31, no. 4, p. 3 - 5, 2009.

ORACLE. Database – What is a Relational Database? Disponível em https://www.oracle.com/database/what-is-a-relational-database/. Acessado em 02 de abril de 2021.

CAFÉ COM BANCO DE DADOS. Banco de Dados Relacional. Disponível em http://www.coffeewithdatabase.com/banco-de-dados-relacional/ Acessado em 02 de abril de 2021.

MICROSOFT AZURE. Dados não relacionais e NoSQL. Disponível em https://docs.microsoft.com/pt-br/azure/architecture/data-guide/big-data/non-relational-data Acessado em 12 de janeiro de 2021.

BANSAL, S. K.; KAGEMANN, S.; Integrating Big Data: A Semantic Extract-Transform-Load Framework. IEEE

HACKERS AND SLACKERS. Learning Apache Spark with PySpark & Databricks. Disponível em https://hackersandslackers.com/learning-to-use-apache-spark-pyspark/ Acessado em 04 de junho de 2021

INFOQ. Pentaho Data Integration - ETL em Software Livre. Disponível em https://www.infoq.com/br/articles/pentaho-pdi/ Acessado em 04 de junho de 2021

WATSON, H.J.; WIXOM, B. H.; The Current State of Business Intelligence. IEEE Computer, Vol. 40, No. 9, 2007

MICROSOFT AZURE. O que são ferramentas de business intelligence (BI)? Disponível em https://azure.microsoft.com/pt-br/overview/what-are-business-intelligence-tools/. Acessado em 04 de junho de 2021

MICROSOFT. 2021 Gartner Magic Quadrant for Analytics and Business Intelligence Platforms. Disponível em https://info.microsoft.com/ww-Landing-2021-Gartner-MQ-for-Analytics-and-Business-Intelligence-Power-BI.html?LCID=EN-US. Acessado em 04 de junho de 2021.

MICROSOFT POWER BI BLOG. Announcing Power BI general availability coming July 24th. Disponível em https://powerbi.microsoft.com/en-us/blog/announcing-power-bi-general-availability-coming-july-24th/. Acessado em 04 de junho de 2021.

MICROSOFT POWER BI. Tutorial: Explorar um exemplo do Power BI. Disponível em https://docs.microsoft.com/pt-br/power-bi/create-reports/sample-tutorial-connect-to-the-samples Acessado em 04 de junho de 2021.

MICROSOFT POWER BI. Tutorial: Introdução à criação no serviço do Power BI. Disponível em https://docs.microsoft.com/pt-br/power-bi/fundamentals/service-get-started Acessado em 04 de junho de 2021.

TERRA. Quem foi o grande responsável e quando surgiu o Excel? Disponível em https://www.terra.com.br/noticias/quem-foi-o-grande-responsavel-e-quando-surgiu-o-excel,f349de3e51654d2473aa2528c45f9e37bmsffnfi.html Acessado em 15 de junho de 2021

NASCIMENTO, J. L.; A utilização do Excel para o ensino de estatística no Ensino Médio: um estudo de caso no município de Mamanguape. UFPB, 2016