

**UNIVERSIDADE FEDERAL DE UBERLÂNDIA
BACHARELADO EM GESTÃO DA INFORMAÇÃO**

DENISE ARAÚJO SANTOS

**DESENVOLVIMENTO DE PROCESSO ETL AUTOMATIZADO EM DADOS DE
PESQUISAS DE UMA EMPRESA NA ÁREA DE TECNOLOGIA**

UBERLÂNDIA – MG

2022

DENISE ARAÚJO SANTOS

**DESENVOLVIMENTO DE PROCESSO ETL AUTOMATIZADO EM DADOS DE
PESQUISAS DE UMA EMPRESA NA ÁREA DE TECNOLOGIA**

Trabalho de conclusão de curso apresentado à Faculdade de Gestão e Negócios da Universidade Federal de Uberlândia, Minas Gerais, como requisito exigido parcial à obtenção do grau de Bacharel em Gestão da Informação.

Orientador: José Eduardo Ferreira Lopes

UBERLÂNDIA – MG

2022

**DESENVOLVIMENTO DE PROCESSO ETL AUTOMATIZADO EM DADOS DE
PESQUISAS DE UMA EMPRESA NA ÁREA DE TECNOLOGIA**

Trabalho de conclusão de curso apresentado à Faculdade de Gestão e Negócios da Universidade Federal de Uberlândia, Minas Gerais, como requisito exigido parcial à obtenção do grau de Bacharel em Gestão da Informação.

Orientador: José Eduardo Ferreira Lopes

Prof. José Eduardo Ferreira Lopes

Prof. Antonio Sérgio Torres Penedo

Prof. Peterson Gandolfi

Uberlândia, 11 de agosto de 2022

RESUMO

O ETL – cuja sigla significa extração, transformação e *load* (carregamento) dos dados – é um processo de integração entre diferentes fontes de dados capaz de gerar *insights* através das possíveis transformações e de armazenar os dados em um único local. Este relato tecnológico tem o objetivo de apresentar a elaboração de um processo ETL automatizado, criado para otimizar atividades realizadas no tratamento de dados de pesquisas, aplicadas por uma empresa da área de tecnologia. Deste modo, propôs-se a utilização da ferramenta Knime por ser intuitiva, rápida para implantação e ser *open-source* (código-aberto), possibilitando a implementação do processo sem custos adicionais. Os resultados obtidos com a aplicação do processo e sua automatização trouxeram aumento na produtividade, melhora na precisão dos dados e na experiência do consumidor, tendo a agilidade como fator principal dos resultados gerando um ganho de 23,8 horas em média a cada execução do processo. Contudo, este relato pode servir de exemplo para implementação em outros projetos e em outras áreas presentes no mercado de trabalho.

Palavras-chave: ETL; Knime; Automatização.

ABSTRACT

ETL – which stands for data extraction, transformation and loading – is an integration process between different data sources capable of generating insights through possible transformations and storing the data in a single location. This technical report aims to describe the steps developed in an automated ETL process, created to optimize activities performed manually in the treatment of research data, applied by a technology company. Thus, it was proposed to use the Knime tool because it is intuitive, quick to implement and is open-source, enabling the implementation of the process without additional costs. The results obtained with the application of the process and its automation brought an increase in productivity, improved data accuracy and consumer experience, with agility as the main factor in the results, generating a gain of 23,8 hours on average for each execution of the process. However, this report can serve as an example for implementation in other projects and in other areas present in the job market.

Keywords: ETL; Knime; Automation.

SUMÁRIO

1 INTRODUÇÃO	7
2 CONTEXTO INVESTIGADO	8
3 DIAGNÓSTICO DA SITUAÇÃO-PROBLEMA	11
4 INTERVENÇÃO PROPOSTA.....	13
5 RESULTADOS OBTIDOS	20
6 CONSIDERAÇÕES FINAIS	20
REFERÊNCIAS BIBLIOGRÁFICAS	22

1 INTRODUÇÃO

Com a crescente escala de dados disponíveis ao longo dos anos, técnicas de manipulação e tratamento de dados, surgiram para auxiliar no processo de análise e coleta de informações para a tomada de decisão no ambiente de negócios, com o objetivo de gerar resultados e vantagem competitiva para as organizações. De acordo com Exame (2021), a cada seis minutos 9,1 mil *terabytes* de dados são gerados no mundo, essa quantidade de dados impõe um grande desafio para os executivos das empresas pela complexidade de interpretar tal volume de informações e muitas das vezes desestruturadas, porém com recursos tecnológicos (ferramentas e aplicações) torna-se possível atender essa necessidade. E dentro desse contexto temos a Inteligência de Negócios, no inglês, *Business Intelligence* (BI).

De acordo com Primak (2008) o termo *Business Intelligence* (BI) surgiu na década de 80 no Garter Group e é composto por uma ampla categoria de aplicações com o objetivo de coletar, organizar, analisar, compartilhar e monitorar dados que gerem informações para auxiliar usuários de negócios nas tomadas de decisões através da implementação de um *Data Warehouse* (DW). O contexto de DW é baseado em um banco de dados que possui diversas fontes com o objetivo de integrar e consolidar os dados, para ser considerado um DW é muito importante que ele possua o processo de *Extract Transform Load* (ETL). O processo de ETL possui a etapa de extração, onde é compilado os dados brutos seja de uma ou mais fontes, a etapa de transformação, onde são realizados tratamentos nos dados para torná-los inteligíveis, por fim, a etapa de carregamento dos dados em um sistema de destino, seja em um DW ou em um banco de dados tradicional, sendo de suma importância para facilitar o acesso e o entendimento dos dados por parte dos tomadores de decisão.

O objetivo deste trabalho é apresentar a elaboração de um processo de ETL automatizado com a ferramenta Knime, criado para otimizar atividades realizadas no tratamento de dados de pesquisas em uma empresa na área de tecnologia, assim como a garantir a qualidade nos dados, sendo esse o principal problema identificado, devido as etapas de transformação dos dados serem executadas por diferentes funcionários e de forma manual, gerando vários erros e inconsistências nos dados ao longo do processo.

A estrutura de apresentação do tema consiste no aprofundamento do contexto investigado, na descrição do diagnóstico da situação problema, da apresentação dos mecanismos adotados para solucionar o problema, na descrição dos resultados obtidos, por fim, as considerações finais.

2 REVISÃO BIBLIOGRÁFICA

De acordo com Kimball e Carseta (2004) um sistema ETL projetado de forma adequada exige padrões de qualidade e consistência de dados, ele é responsável por 70% dos recursos necessários para implementação de um DW, sendo crucial para definir um DW consistente. Além disso, ressaltam que o sistema ETL é muito mais que uma estrutura para obter dados de diversas fontes e inserir no DW, ele agrega valor significativo aos dados removendo erros, corrigindo dados ausentes, ajustando os dados de várias fontes para serem usados juntos, sendo adequável as regras de negócios das organizações.

Para Abreu (2008), o processo de ETL não é apenas um subprocesso para a construção de um DW ou parte de um projeto de BI, mas sim essencial para toda estrutura de sistemas OLTP (*On-Line Transaction Processing* ou Processamento de Transações em Tempo-Real) transacionais para otimização dos processos de TI nas organizações.

O processo de ETL é composto por três etapas: extração, transformação e carga dos dados, ilustradas na Figura 1. As etapas de extração e carga são obrigatórias acontecerem no processo, sendo a transformação uma etapa opcional.

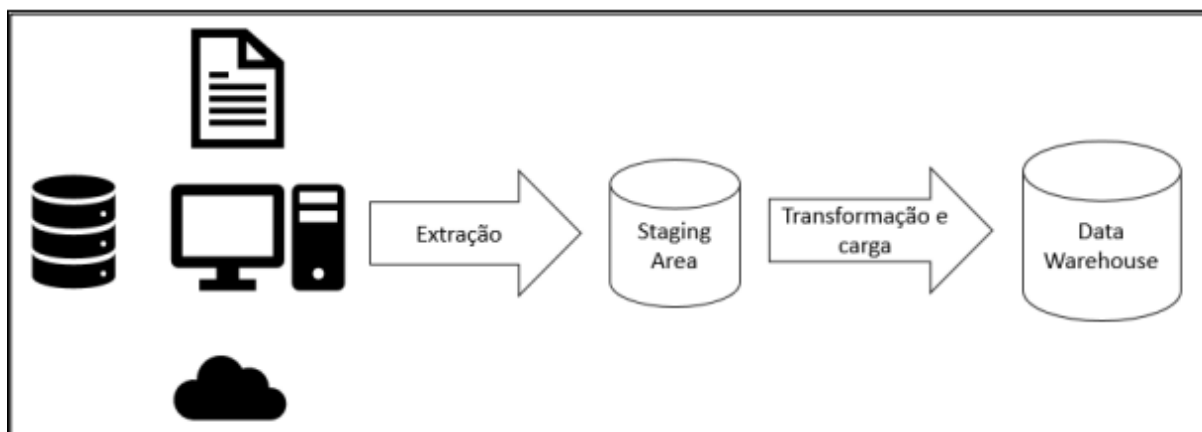


Figura 1. Estrutura do processo de ETL
Fonte: Adaptado de Kimball e Caserta (2004)

Na etapa de extração é onde ocorre a integração entre sistemas (fontes) diferentes sendo este o maior desafio enfrentado pelas organizações e o primeiro passo para enfrentar este desafio é a construção de um mapa lógico de dados que contém todas as especificações funcionais para construir um processo de ETL (KIMBALL; CARSETA, 2004). Já para Liu e Özsü (2009) o processo de identificação do subconjunto correto de dados de origem a ser inserido no fluxo de trabalho do processo de ETL é considerado uma tarefa simples.

Segundo Lane e Schupmann (2005) a escolha do método de extração está ligada ao sistema de origem e as regras do negócio e nem sempre há possibilidade de implementar lógica na etapa de extração pois pode haver o comprometimento do desempenho e aumento de cargas de trabalho no processo. Ainda segundo Lane e Schupmann (2005), há dois tipos de métodos de extração de dados, o método de extração lógica e o de extração física.

- O método de extração lógica pode ser completo ou incremental, na extração completa os dados são totalmente extraídos do sistema de origem sem o controle de alterações desde a última extração com sucesso. Já na incremental a extração ocorre somente com dados que foram alterados ou dados que são novos.
- O método de extração física pode ser *online* ou *offline*, a extração *online* é realizada diretamente na fonte dos dados ao contrário da *offline* onde a extração é feita de uma fonte externa que possui cópia dos dados de origem.

A segunda etapa do processo de ETL é a transformação sendo considerada a etapa mais cara do processo, pois variam de simples conversões a técnicas de depuração de dados complexas (LANE; SCHUPMANN, 2005). Já para Kimball e Caserta (2004) a etapa de transformação se resume a limpeza e conformidade dos dados que geram metadados potentes ao processo, pois eles evidenciam várias técnicas e métodos para tratativas de possíveis problemas já na etapa de extração.

A transformação dos dados é possível através de diversas funções e regras de ajustes sendo algumas delas: alteração de coluna, alteração de registro, junção de tabelas e criação de novos registros a partir de outros. E como dito anteriormente, ela é opcional no processo de ETL, um processo sem a etapa transformação tem a utilidade de transportar os dados de um sistema de origem para um sistema de destino.

A última etapa do processo de ETL é o carregamento dos dados, segundo Ponniah (2011) a etapa de carregamento pode levar muito tempo, gerando preocupação em relação a boa eficiência desta etapa, o processo deve ser capaz de identificar erros na inserção e reportá-los em um *log*. Ainda de acordo com Ponniah (2011) existem três tipos de *load* (carregamento) de dados:

- **Load Inicial:** ocorre quando os dados são inseridos no banco de dados pela primeira vez;
- **Load Incremental:** insere somente dados atualizados, sendo uma forma de identificar a última vez que os dados foram alterados;
- **Full Load:** ocorre quando os dados do sistema de destino são apagados para serem carregados todos os dados do sistema de origem.

As ferramentas disponíveis no mercado atualmente são capazes de processar milhões de dados em segundos, também é possível realizar trabalhos que levariam um mês, em horas com as ferramentas de ETL e há uma grande variedade disponível (TAVARES, 2013). Algumas das ferramentas citadas por Tavares (2013) foram IBM Information Server, Talend Studio for Data Integration, SAS Data Integration Studio, Pentaho, QlikTech, entre outras.

Segundo Oliveira (2019) o Talend Open Studio é ideal para integração de dados entre sistemas, onde o fluxo de trabalho é representado pela modelagem gráfica de processos podendo ser usado tanto por programadores quanto por pessoas que não possuem essa habilidade.

De acordo com Knime (2022), a ferramenta utilizada neste estudo, é uma plataforma *open-source* (código-aberto) que auxilia na descoberta de potencial oculto nos dados, exploração de novos insights e previsão de recursos. Sendo rápida para implantação e intuitiva para aprender. Na Knime Analytics Platform, as tarefas a serem realizadas no fluxo de trabalho são representadas por nós, uma série de nós interconectados dão origem ao um processo de ETL. Cada nó possui portas de entradas e saídas, as de entrada armazenam os dados que o nó é responsável por processar e as de saída armazenam os conjuntos de dados resultantes da operação, elas só podem ser conectadas a uma porta de entrada do mesmo tipo, como mostra a Figura 2.

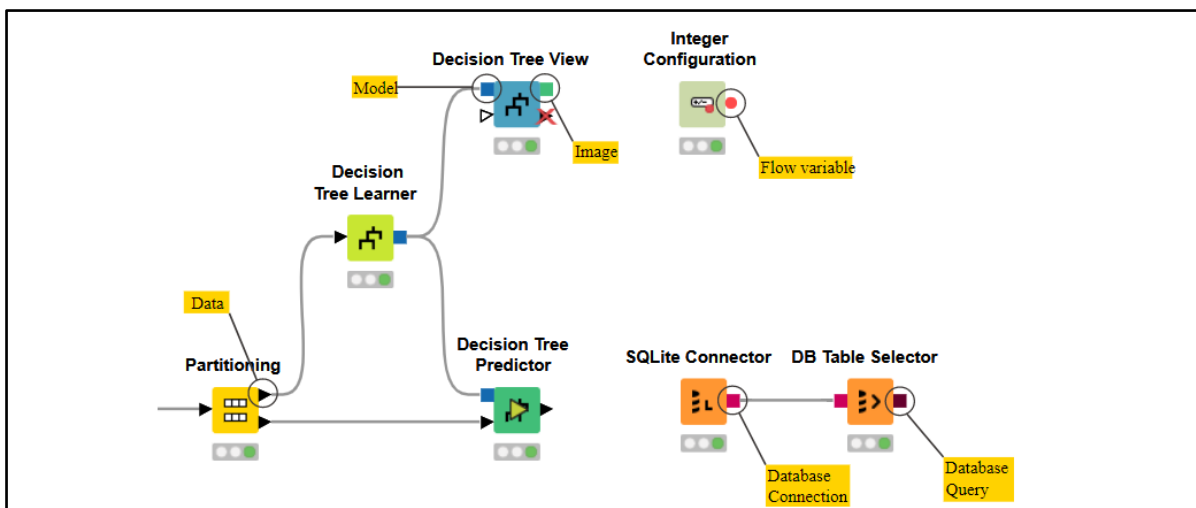


Figura 2. Tipos de portas e suas conexões
Fonte: Guia do Knime Workbench (2021)

A automatização de processos ao contrário do que as pessoas imaginam, é antiga e foi evidenciada a partir do século XVII pela Revolução Industrial através da mecanização (TEZA, 2002). Ainda segundo Teza (2002), automatização é definida como qualquer processo que facilite as tarefas rotineiras para o ser humano com a utilização de dispositivos automáticos para tal. Um exemplo de automatização pode-se dar com um aparelho doméstico considerado essencial, a geladeira, nela existem mecanismos e tecnologias que permitem manter os alimentos gelados constantemente de forma automática. Diante disso, pode-se dizer que a automatização é muito importante para a obtenção dos resultados deste estudo.

3 DIAGNÓSTICO DA SITUAÇÃO-PROBLEMA

O presente trabalho foi elaborado através de uma análise no contexto de negócios, avaliando as atividades relacionadas ao projeto de pesquisa de uma empresa que presta esse serviço ao seu cliente, ambos no segmento de tecnologia. Tanto a prestadora do serviço de pesquisa, quanto seu cliente não serão identificados neste trabalho por questões de sigilo e segurança da informação, sendo assim, a empresa prestadora de serviços é citada no texto como Empresa A.

A Empresa A, possui soluções voltadas para a transformação digital dos processos de vendas e relacionamento de seus clientes e está há 12 anos gerando resultados para seus clientes através da fusão de produtos estratégicos para o mercado. As pesquisas são realizadas tanto por telefone quanto presenciais, porém

nos últimos anos devido a pandemia do COVID-19 as pesquisas passaram a ser somente por telefone. Os recursos necessários para a realização das pesquisas variam de acordo com a necessidade do cliente, sendo assim, toda pesquisa é tratada como um projeto e são aplicadas técnicas de gestão para se obter uma entrega bem-sucedida dos resultados.

De acordo com Cruz (2013) um projeto é um esforço temporário aplicado seja para criação de um produto, serviço ou atingir um resultado. Dentro deste contexto, a equipe de negócios fica responsável pela execução dos projetos de pesquisas da empresa, suas atividades englobam acompanhar o andamento das etapas a serem cumpridas, fazer acompanhamento dos prazos de entregas determinados no *Onboarding* (processo de integração entre as pessoas envolvidas no projeto), além de monitorar a coleta das informações de pesquisas, analisar os resultados obtidos e apresentar ao cliente.

O problema identificado foi entre as atividades de coletas e análises dos resultados, como se trata de um projeto que conta com várias pessoas envolvidas, as atividades são executadas praticamente de forma simultânea, na etapa de coleta especificamente, acontecem entregas parciais com amostras até o montante final de respostas, todas as amostras de respostas passam por um processo de tratamento para ser possível analisá-las. O tratamento inclui extração da base de respostas, manipulação de *strings* (sequência de caracteres que formam dados textuais), limpeza de respostas inválidas e organização de colunas, este é feito de forma manual com auxílio da ferramenta Excel (ferramenta de edição de planilhas) o que demanda tempo e esforço humano com alto risco de haver erros e inconsistências nos dados. Segundo Oliveira (2020) algumas características como verificações à “olho nu” e falta de automatização no tratamento de dados podem fomentar o chamado “ETL Humano”, uma cultura indesejada que cria silos de dados e inconsistências nas informações das empresas. O fluxo de trabalho manual e suas etapas são apresentados na Figura 3.

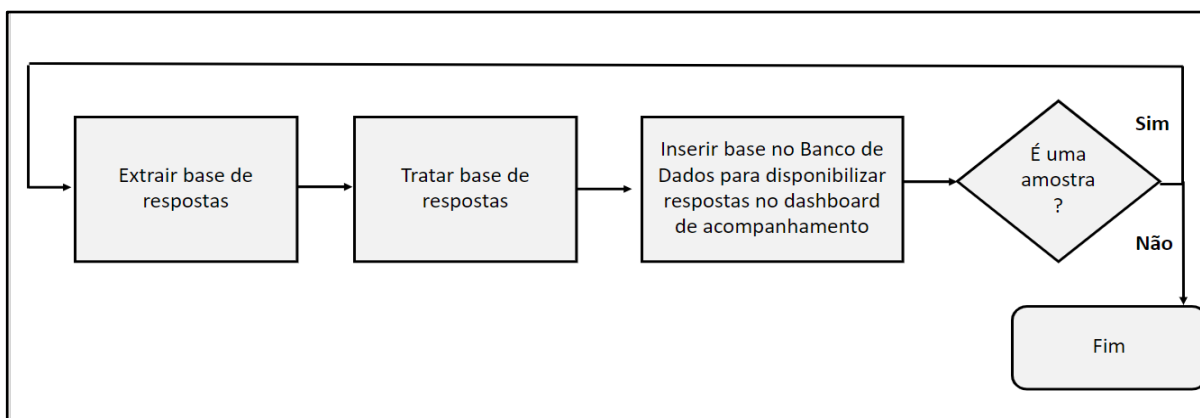


Figura 3. Fluxo de trabalho manual.

Fonte: Autora (2022)

A partir do problema identificado chegou-se a seguinte questão: Como otimizar as atividades de tratamento dos dados oriundos das pesquisas? No próximo capítulo será apresentada a solução implantada para tal.

4 INTERVENÇÃO PROPOSTA

A pesquisa utilizada como objeto de estudo na análise citada anteriormente, trata-se de uma avaliação de perda e ganho em relação aos produtos oferecidos pelo cliente da Empresa A, ela ocorre mensalmente, sendo uma pesquisa recorrente. Com a pesquisa é possível mapear de onde as pessoas vieram, se analisaram propostas de concorrentes, nível de satisfação, para onde foram no caso das que recusaram os produtos oferecidos e perfil dos decisores. Todos os dados da pesquisa são protegidos pela LGPD (Lei de Proteção Geral dos Dados) onde os respondentes são identificados mediante autorização dos mesmos.

O processo criado com a utilização da ferramenta Knime foi composto por 29 componentes abrangendo interconexões de nós e metanodes de vários tipos para contemplar as etapas de extração, transformação e carregamento dos dados presentes na execução das atividades de forma manual, como visto anteriormente. Na Figura 4, temos a visão geral do projeto de tratamento na ferramenta Knime e ao longo do texto suas partes serão especificadas.

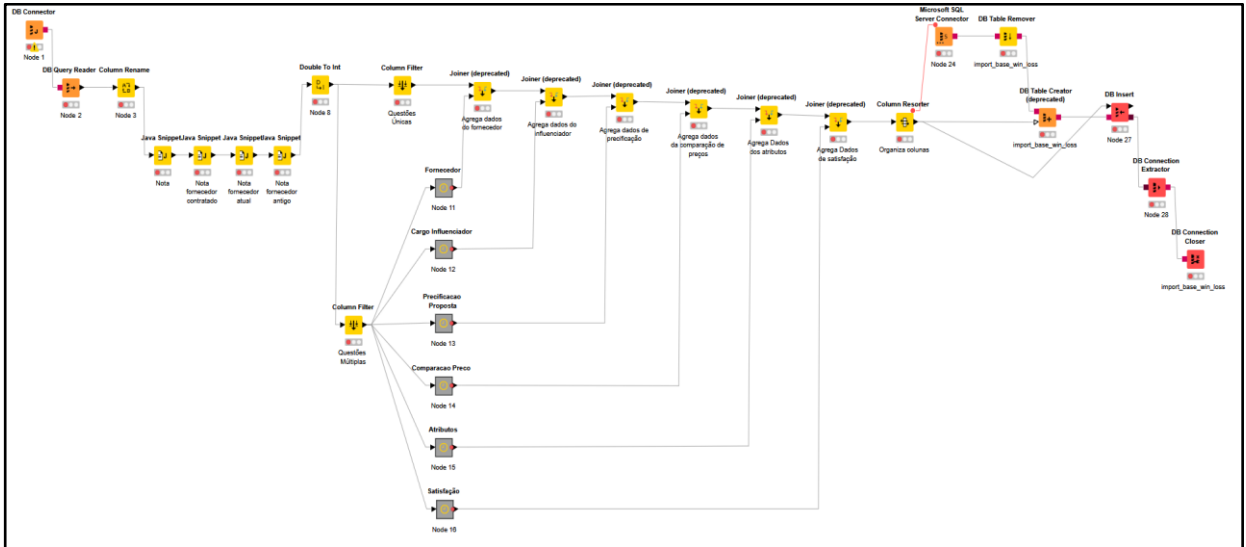


Figura 4. Projeto de tratamento.
Fonte: Knime Analytics Platform (2022)

O primeiro nó DB Connector, é responsável pela conexão com o banco de dados onde estão armazenadas as bases com as perguntas e respostas da pesquisa, na sequência o nó DB Query Reader é responsável por ler estes dados e no Column Rename é realizada a codificação das perguntas através da renomeação das colunas, esse passo é importante para padronizar os nomes das colunas e garantir que não haja presença de caracteres especiais que impeçam a carga da base tratada no banco de dados. Os nós citados foram evidenciados na Figura 5.

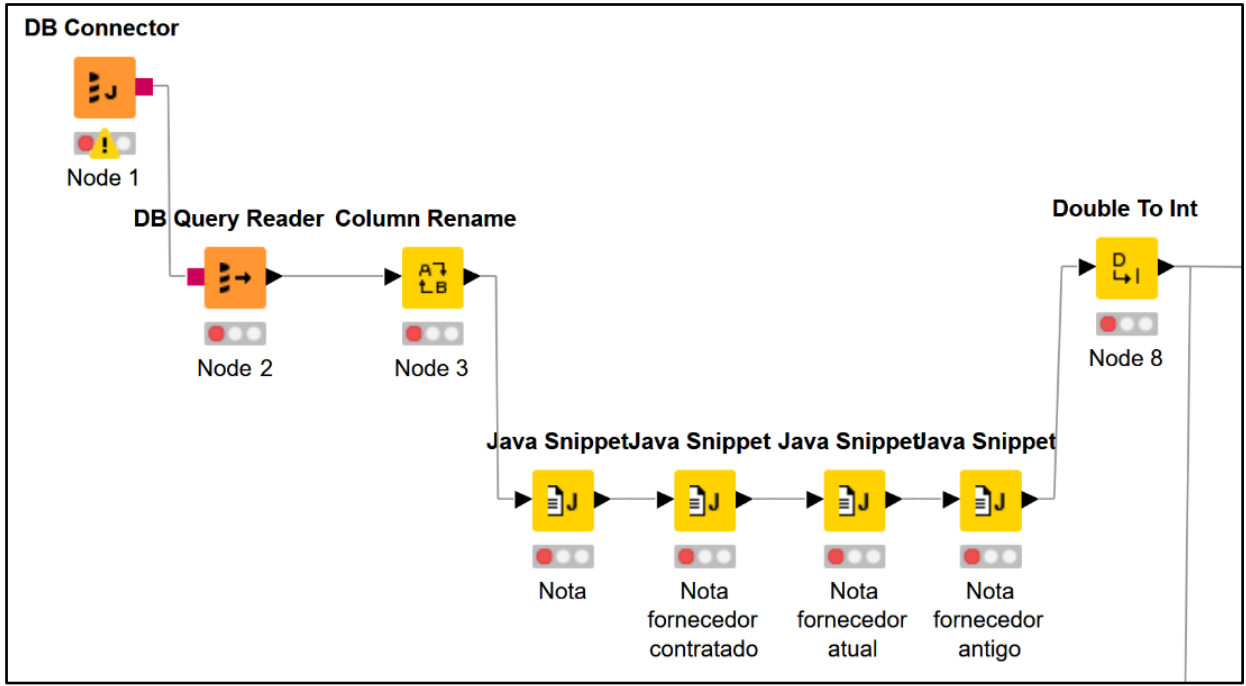


Figura 5. Projeto de tratamento.
Fonte: Knime Analytics Platform (2022)

Acompanhando o fluxo do processo de tratamento da Figura 5, há uma sequência de nós Java Snippet que são responsáveis por permitir a manipulação de dados através da linguagem Java, todos executam a mesma função, porém para colunas diferentes. Sua função é a de alterar as respostas em branco ou marcadas como “Não sabe” para nulo, pois inicialmente essas respostas são classificadas como *strings* e não entram no cálculo da métrica de avaliação da satisfação dos respondentes. A métrica é calculada para as notas que os respondentes deram de 1 a 10 para o cliente da Empresa A, caso o respondente não tenha adquirido seus produtos, ele avalia o fornecedor contratado e caso ele tenha mudado de fornecedor, avalia o atual e o antigo. O nó de número 8 é o Double To Int responsável por transformar números decimais em números inteiros e essa parte do processo garante que todos os dados numéricos da base sejam do mesmo tipo.

Na próxima etapa do processo é feito o tratamento das questões de múltipla escolha, elas necessitam de tratamento pois cada opção escolhida pelo respondente é armazenada em colunas diferentes, não sendo viável a visualização das respostas no *dashboard* de acompanhamento onde ele mostra todas as repostas para cada pergunta do questionário, o formato ideal seria cada pergunta ser representada por uma coluna com suas respectivas respostas em linhas. Sendo assim, há dois nós Column Filter na sequência responsáveis por filtrar as colunas, um filtra somente as colunas que se referem às questões com respostas únicas em uma base de dados e o outro filtra as colunas que se referem às questões com respostas de múltipla escolha noutra base de dados, como mostra a Figura 6.

Após a filtragem, cada pergunta de múltipla escolha foi tratada em um nó chamado Metanode, ele permite que um novo fluxo de tratamento seja criado dentro dele, neste caso foi usado um fluxo de tratamento padrão para cada questão de múltipla escolha, o questionário é composto por seis perguntas de múltipla escolha sendo elas sobre atributos do fornecedor, do perfil do influenciador, do preço oferecido, da comparação de preço com outras propostas, do cliente e da satisfação em relação ao atendimento prestado, prazo de envio da proposta, clareza da proposta, inovação e preço, como mostra a Figura 6.

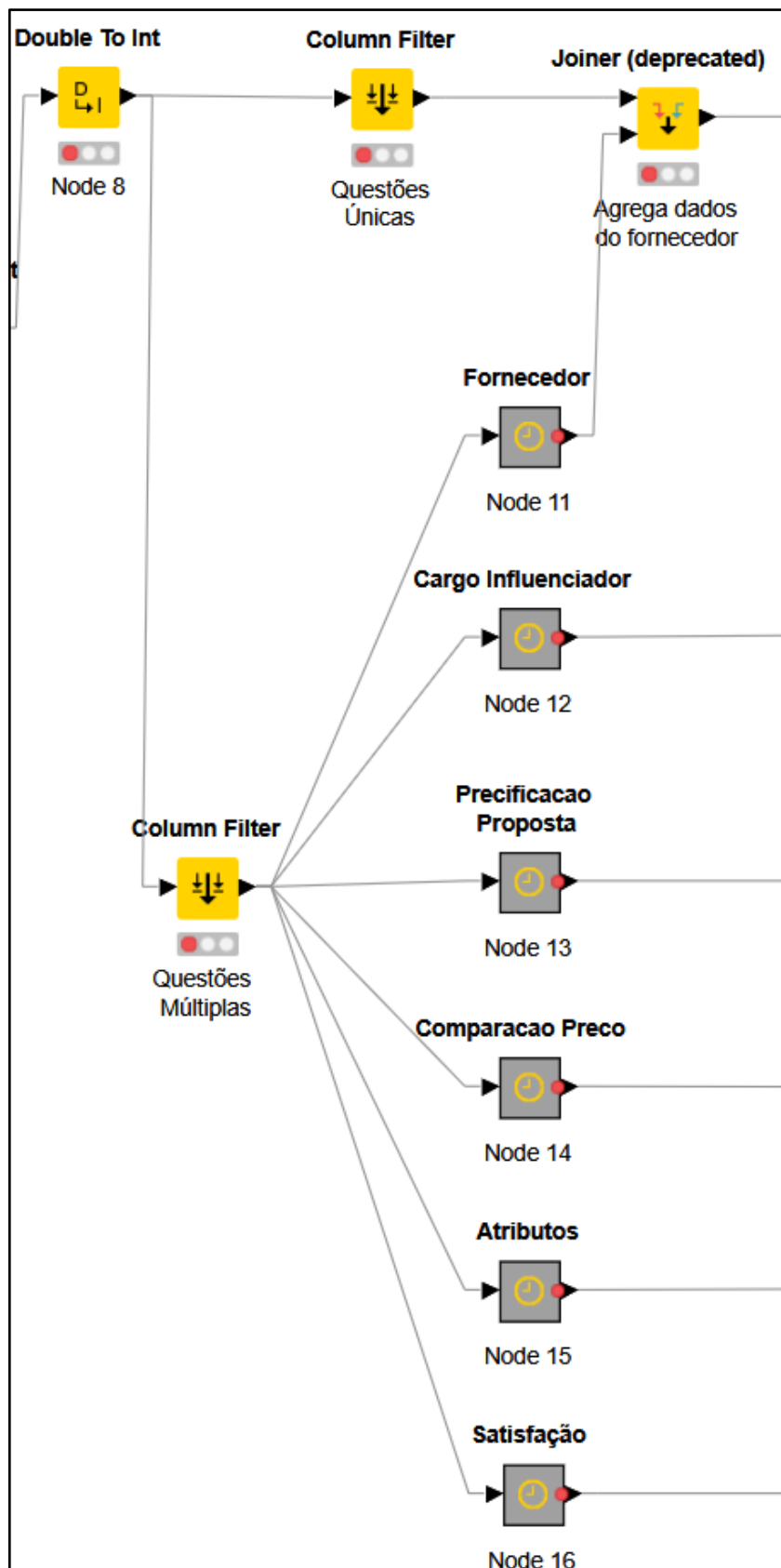


Figura 6. Projeto de tratamento.
Fonte: Knime Analytics Platform (2022)

Dentro de cada Metanode há um fluxo de tratamento padrão contendo um nó Column Filter para filtrar as colunas correspondentes a pergunta descrita no Metanode, um nó chamado Unpivoting responsável por transformar colunas em linhas, deixando todas as respostas que antes estavam separadas por colunas dispostas em linhas, tendo como base o respondente id (identificador único de cada respondente) para atribuir todas as respostas ao respondente correto, um nó Column Rename para codificar a nova coluna criada no Unpivoting e por fim mais um nó Column Filter para selecionar somente a coluna codificada e o respondente id que são as informações que serão agregadas à base que contém as questões únicas. Após a execução desse fluxo em todos os Metanodes, seis bases são geradas cada uma contendo respostas em colunas únicas e o respondente id, essas informações são agregadas à base com questões únicas através do nó chamado Joiner, ele é responsável por unir informações que estão em duas bases diferentes por meio de um identificador único que neste caso é o respondente id. O Joiner possui duas entradas e uma saída que é o resultado da junção das informações, sendo assim, foi necessário aplicar um Joiner para cada junção como mostra a Figura 7.

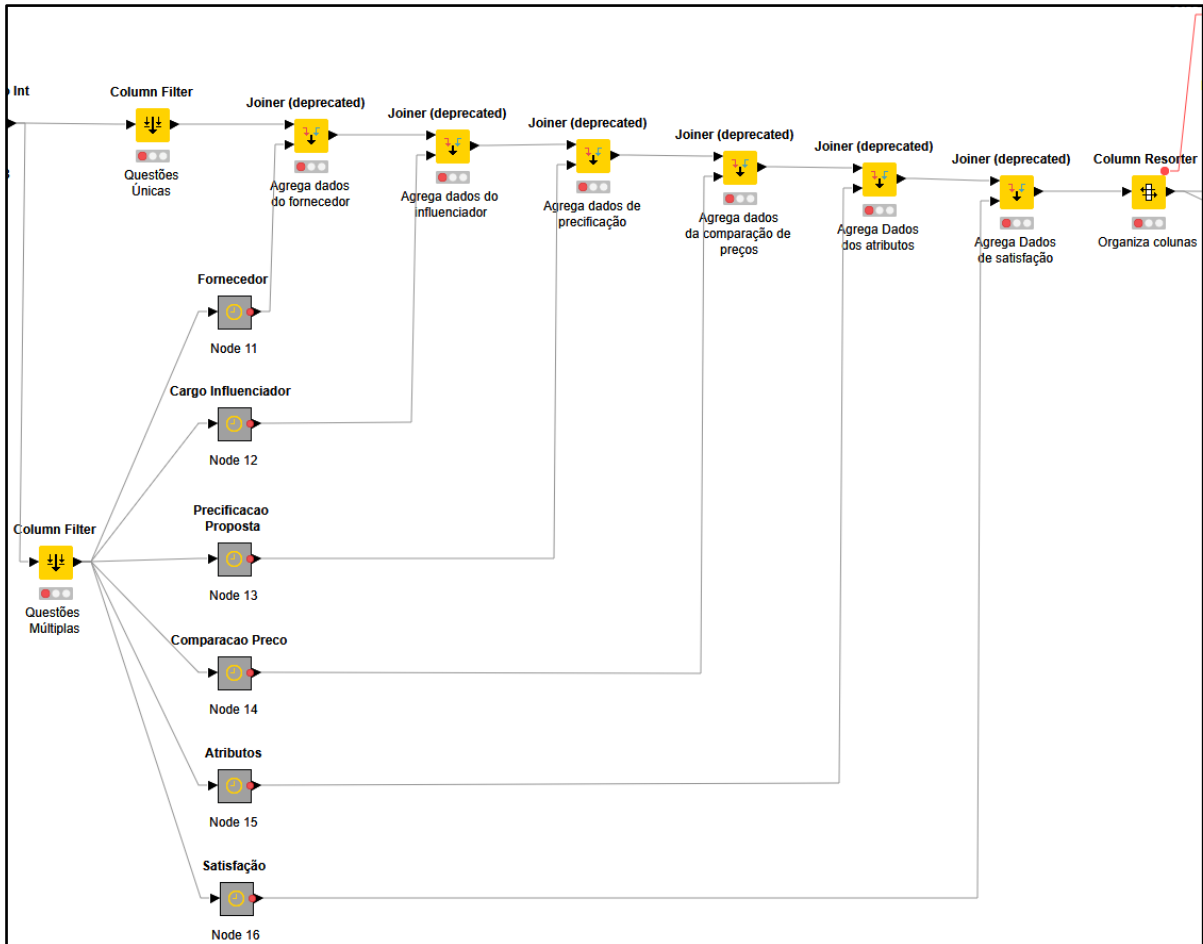


Figura 7. Projeto de tratamento.
Fonte: Knime Analytics Platform (2022)

Em seguida o nó Column Resorter é responsável por ordenar as colunas e permite que sejam organizadas conforme preferência, neste caso elas foram organizadas seguindo a ordem das perguntas no questionário, sendo esse o último processo de tratamento realizado. Após este processo, vem a etapa de carga no banco de dados que disponibiliza as informações no dashboard de acompanhamento. Ela é composta pelos nós Microsoft SQL Server Connector responsável pela conexão com o banco, DB Table Remover responsável por excluir os dados que foram gravados anteriormente, DB Table Creator responsável por criar uma nova tabela no banco, sendo um espaço criado para receber os novos dados, DB Insert responsável por inserir os dados atualizados na tabela criada no nó anterior, DB Connection Extrator responsável por extrair informações das transações realizadas no banco enquanto a conexão está aberta e o DB Connection Closer responsável por fechar a conexão com o banco de dados, esse nó é importante para garantir que a conexão

não fique ociosa toda vez que o processo executar. Os nós citados foram evidenciados na Figura 8.

O projeto de tratamento, chamado como *job* foi programado para executar todo dia 10 até o dia 21 de cada mês que são os dias programados nas *sprints* do projeto de pesquisa para iniciar a coleta das amostras e finalizar a aplicação do questionário. O *job* é executado de forma automática sem a necessidade de nenhuma intervenção humana, todas as etapas que eram feitas de forma manual para ser possível analisar os dados da pesquisa são realizadas em minutos, salvo que só será necessária alguma intervenção caso algum padrão programado mude, por exemplo, alteração ou inclusão de alguma pergunta no questionário, mudanças nas credenciais de conexão com os bancos, entre outras.

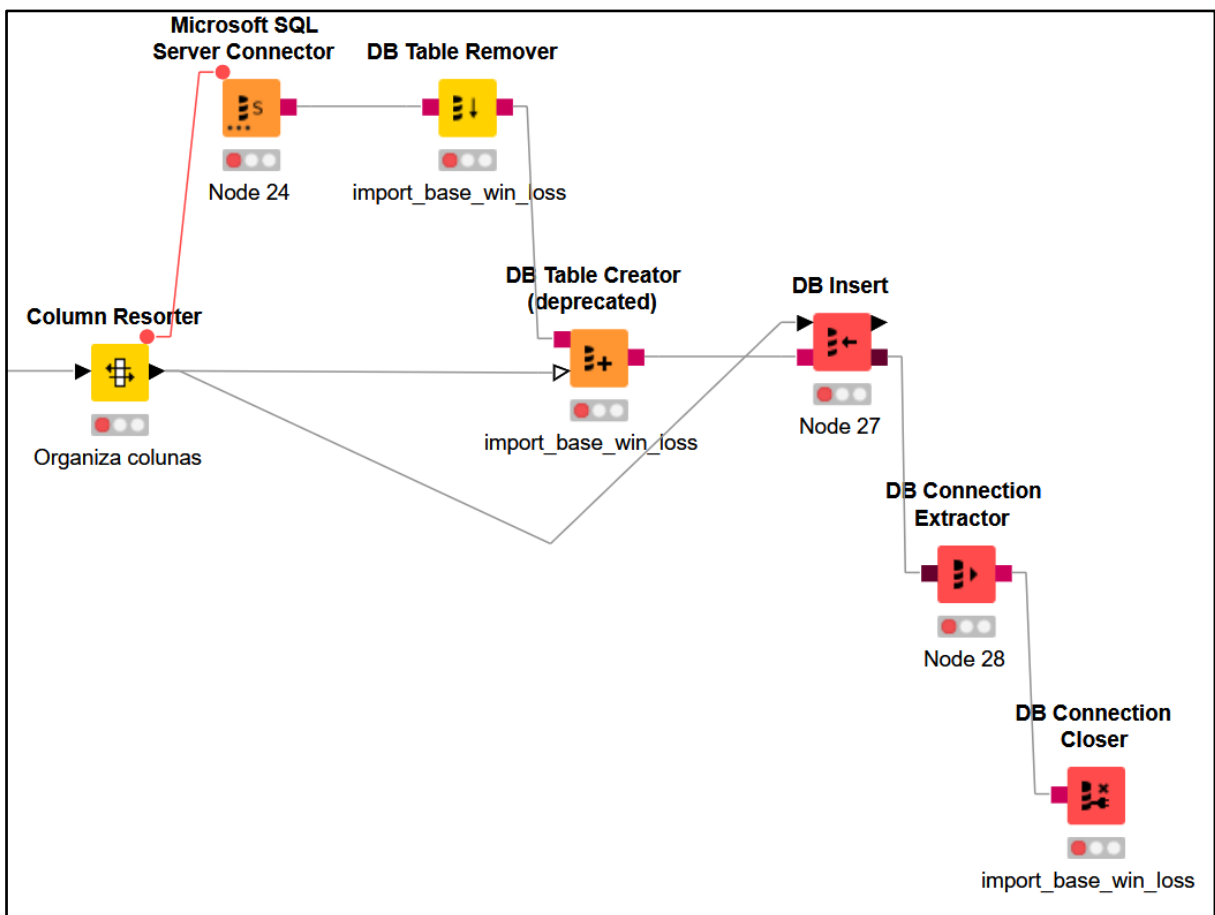


Figura 8. Projeto de tratamento.
Fonte: Knime Analytics Platform (2022)

5 RESULTADOS OBTIDOS

A aplicação da solução trouxe resultados relacionados ao aumento de 11,5% no índice de produtividade, aumento de 3,6% na margem bruta do projeto da pesquisa usada neste estudo, maior precisão nos dados da pesquisa e melhora na experiência do consumidor. O fator principal para esses indicadores é o tempo de execução, pois antes da implementação da solução, o analista gastava em média 24 horas (3 dias trabalhados) para concluir o tratamento de uma amostra de respostas e toda amostra extraída tinha que passar pelos mesmos processos de tratamentos, sendo 24 horas para cada execução. Após o fim da aplicação do questionário previsto para todo dia 20 de cada mês o analista realizava o último tratamento gerando a base de respostas final e tinha em média 7 dias trabalhados para analisar os dados, montar apresentação dos resultados e apresentá-los ao cliente, a apresentação para o cliente geralmente não cumpria o prazo proposto na *sprint* do projeto e ocorria sempre no início do mês posterior, onde havia o acúmulo de tarefas pois já estavam previstas outras atividades de início da pesquisa do mês corrente.

Após a aplicação da solução o tempo para a execução das atividades foi otimizado em 75% com um ganho médio de 23,8 horas por execução, sendo possível analisar os dados enquanto o questionário é aplicado. O analista consegue realizar a apresentação dos resultados da pesquisa 2 dias após o fim da aplicação do questionário, ou seja, no dia 22 de cada mês finalizando a última *sprint* 8 dias antes do previsto. Essa solução impactou positivamente a margem bruta do projeto pois são previstas menos horas para conclusão das atividades e não foi preciso investir na ferramenta utilizada, a precisão dos dados pois não há o risco de erros e silos nos dados durante o processo de tratamento garantindo consistência e confiabilidade nos dados e a satisfação do cliente da Empresa A que agora tem acesso aos dados das pesquisas com antecedência e precisão contribuindo para sua vantagem competitiva.

6 CONSIDERAÇÕES FINAIS

O estudo apresentado gerou crescimento nos resultados da empresa tanto no aspecto financeiro com aumento de margem bruta, como no aspecto operacional com aumento da produtividade, cumprimento dos prazos estipulados e aumento no

indicador NPS, servindo como ponto de partida para outras melhorias que podem ser feitas utilizando a mesma ferramenta e o mesmo modelo de tratamento, sendo uma inspiração para que a Empresa A otimize outras atividades relacionadas a pesquisas de outros clientes aumentando sua performance e vantagem competitiva no mercado. A apresentação da ferramenta Knime é uma breve amostra do que é possível fazer com ela, salvo que ela possui inúmeras funções e nós para vários tipos de problemas a serem resolvidos, pode servir também para que outras empresas conheçam e apliquem modelos automatizados para otimização de atividades sem a necessidade de alto investimento.

Uma sugestão para que a solução aplicada se torne escalável para a Empresa A é a padronização de processos desde a elaboração do questionário até o resultado das pesquisas e também desenvolver um modelo automatizado de tratamento para os *mailings* (lista de contatos) das pesquisas que ainda são tratados manualmente.

Contudo o estudo foi de grande valia para o aprendizado desta autora com aprofundamento na área de tecnologia especificamente em dados, automação e negócios.

REFERÊNCIAS BIBLIOGRÁFICAS

ABREU, Fábio S. G. G e. **Desmistificando o conceito de ETL**. Faculdade Salesiana Maria Auxiliadora, Macaé, Dez.2008. Disponível em <http://www.fsma.edu.br/si/Artigos/V2_Artigo1.pdf> Acesso em: 21 jul.2022.

CRUZ, Fábio. **Scrum e PMBOK unidos no Gerenciamento de Projetos**. Brasport, 2013.

EXAME. **Temos mais dados do que nunca. Como usá-los a nosso favor?** Jun. 2021. Disponível em <<https://exame.com/carreira/dados-uso-favor/>> Acesso em: 11 mar.2022

FÁTIMA, Nilda. **Processo ETL e as etapas para sua implementação**. Fev.2020. Disponível em <<https://www.astera.com/pt/tipo/blog/etl-processo-e-etapas/>> Acesso em: 28 fev.2022.

KIMBALL, Ralph; CASERTA, Joe. **The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data**. Indianapolis: Wiley Publishing, Inc., 2004.

KNIME A. G. **KNIME Workbench Guide**, Zurich, Switzerland, Version 4.6. Abr.2022. Disponível em <https://docs.knime.com/latest/analytics_platform_workbench_guide/index.html#workspaces> Acesso em: 21 jul.2022.

LANE, Paul; SCHUPMANN, Viv. **Oracle Database Data Warehousing Guide**, 11g Release 2 (11.2) E16579-01, 2005.

LIU, Ling; ÖZSU, M. Tamer (Ed.). **Encyclopedia of database systems**. New York, NY, USA, Springer, 2009.

PONNIAH, Paulraj. **Data warehousing fundamentals for IT professionals**. John Wiley & Sons, 2011.

PRIMAK, Fábio Vinícius. **Decisões com bi (business intelligence)**. Fabio Vinicius Primak, 2008.

TAVARES, Edmir J. O. **Processo ETL: O caso da Unitel T+ Telecomunicações**. Universidade Jean Piaget de Cabo Verde, Cidade da Praia, 2013.

TEZA, V. R. **Alguns Aspectos Sobre a Automação Residencial – Domótica**. Universidade Federal de Santa Catarina – Programa de Pós-Graduação em Ciência da Computação. Florianópolis, 2002.

OLIVEIRA, Grimaldo. **ETL Humano - Sua Empresa Possui?** Jun. 2020. Disponível em <<https://www.linkedin.com/pulse/etl-humano-sua-empresa-possui-msc-grimaldo-oliveira>> Acesso em: 28 fev.2022.

OLIVEIRA, Marcio Rodrigues de. **ETL - Overview da ferramenta Talend Open Studio – Data Integration**. Nov. 2019. Disponível em <<https://pt.linkedin.com/pulse/etl-overview-da-ferramenta-talend-open-studio-data-marcio>> Acesso em: 22 jul.2022.