



UNIVERSIDADE FEDERAL DE UBERLÂNDIA  
Faculdade de Gestão e Negócios – FAGEN  
Bacharelado em Gestão da Informação

JOÃO PAULO GUIMARÃES FLORES MORAIS

**A USABILIDADE DO T-SQL NA CONSTRUÇÃO E MANUTENÇÃO DE PROCESSOS  
ETL**

Uberlândia – MG

2023



---

JOÃO PAULO GUIMARÃES FLORES MORAIS

**A USABILIDADE DO T-SQL NA CONSTRUÇÃO E MANUTENÇÃO DE PROCESSOS  
ETL**

Trabalho de Conclusão de Curso apresentado à Banca Examinadora como requisito parcial para obtenção do título de Bacharel em Gestão da Informação da Faculdade de Gestão e Negócios – Universidade Federal de Uberlândia.

***Orientador:** Prof. Dr. José Eduardo Ferreira Lopes*

Aprovado por:

---

Prof.<sup>a</sup> Dr. José Eduardo Ferreira Lopes  
Faculdade de Gestão e Negócios - FAGEN  
**ORIENTADOR**

---

Prof. Dr. Peterson Elizandro Gandolfi  
Faculdade de Gestão e Negócios – FAGEN

---

Prof. Dr. Antônio Sérgio Torres Penedo  
Faculdade de Gestão e Negócios - FAGEN

---

Doutorando Denílson Carrijo Ferreira  
Faculdade de Gestão e Negócios – FAGEN

Uberlândia, 22 de junho de 2023



---

## RESUMO

O presente relato tecnológico tem como objetivo apresentar a usabilidade da linguagem T-SQL (*Transact-SQL*) - uma variação da linguagem SQL (*Structured Query Language*) desenvolvida para interagir com o banco de dados *Microsoft SQL Server* - na construção e manutenção de processos ETL (*Extract, Transform and Load*, do português: Extração, Transformação e Carga) nos ambientes de *Business Intelligence* e *Business Analytics*. No relato tecnológico em questão, realizamos uma revisão da bibliografia para compreender as principais características e funcionalidades que versam os conceitos de *Business Intelligence*, *Business Analytics*, processos ETL e da linguagem T-SQL, explorando a adequação dos referidos conceitos para a implementação e manutenção de processos ETL. Deste modo, valendo-se da utilização do T-SQL e de seu ecossistema *Microsoft*, apresentamos a construção e manutenção dos Processos ETL de uma empresa brasileira líder em um setor estratégico, onde, os resultados obtidos nos permitiram compreender os desafios enfrentados pelos desenvolvedores e analistas, as melhores práticas adotadas por eles e os benefícios obtidos com o uso da linguagem neste contexto, onde, constatou-se que o T-SQL é uma linguagem poderosa e flexível para a construção e manutenção de processos ETL. No entanto, também identificamos algumas limitações da linguagem em questão aplicada aos processos ETL, principalmente no que tange a complexidade de operacionalização, manutenção e custos dos referidos processos.

**Palavras-Chave:** *Analytics. Banco de Dados. Business Intelligence. ETL. Microsoft SQL Server. T-SQL.*



---

## ABSTRACT

This technological report aims to present the usability of the T-SQL language (Transact-SQL) - a variation of the SQL language (Structured Query Language) developed to interact with the Microsoft SQL Server database - in the construction and maintenance of ETL processes (Extract, Transform and Load, from Portuguese: Extraction, Transformation and Load) in Business Intelligence and Business Analytics environments. In this technological report, we carry out a review of the bibliography to understand the main characteristics and functionalities that deal with the concepts of Business Intelligence, Business Analytics, ETL processes and the T-SQL language, exploring the suitability of these concepts for the implementation and maintenance of ETL processes. Thus, using T-SQL and its Microsoft ecosystem, we present the construction and maintenance of the ETL Processes of a leading Brazilian company in a strategic sector, where the results obtained allowed us to understand the challenges faced by developers and analysts, the best practices adopted by them and the benefits obtained from using the language in this context, where it was found that T-SQL is a powerful and flexible language for building and maintaining ETL processes. However, we also identified some limitations of the language in question applied to ETL processes, mainly regarding the complexity of operationalization, maintenance and costs of those processes.

**Keywords:** *Analytics. Business Intelligence. Database Management ETL. Microsoft SQL Server. T-SQL.*



---

## LISTA DE SIGLAS

<b>BA</b>	<i>Business Analytics</i>
<b>BI</b>	<i>Business Intelligence</i>
<b>BPM</b>	<i>Business Process Modeling Notation</i>
<b>DER</b>	<i>Diagrama Entidade Relacionamento</i>
<b>DW</b>	<i>Data Warehouse</i>
<b>ERP</b>	<i>Enterprise Resource Planning</i>
<b>ETL</b>	<i>Extract, Transform e Load</i>
<b>KPI'S</b>	<i>Key Performance Indicators</i>
<b>SP</b>	<i>Stored Procedure</i>
<b>SSMS</b>	<i>SQL Server Management Studio</i>
<b>T-SQL</b>	<i>Transact-SQL</i>



---

## SUMÁRIO

1. INTRODUÇÃO.....	7
2. REFERENCIAL TEÓRICO.....	11
2.1. Business Intelligence e Business Analytics.....	11
2.1.1. Business Intelligence.....	12
2.1.2. Business Analytics.....	13
2.2. ETL.....	15
2.3. Transact-SQL (T-SQL).....	17
3. SITUAÇÃO PROBLEMA.....	18
4. INTERVENÇÃO ADOTADA.....	19
4.1. Objetivos.....	19
4.2. Metodologia.....	20
4.3. Desenvolvimento.....	25
5. RESULTADOS OBTIDOS.....	30
6. CONSIDERAÇÃO FINAIS.....	31
REFERÊNCIAS BIBLIOGRÁFICAS.....	32



---

## 1. INTRODUÇÃO

A capacidade de extrair, transformar e carregar dados com fluidez e segurança é um componente fundamental para a obtenção de *insights* relevantes nos ambientes de *Business Intelligence* e *Business Analytics*. Os processos ETL (Extração, Transformação e Carga) desempenham um papel crucial neste contexto, permitindo a integração de dados oriundos de diversas fontes, sua transformação em formatos adequados e a carga no local de armazenamento apropriado e previamente adequado ao uso (KIMBALL; CASERTA, 2004). A partir disto, os dados em questão podem ser explorados e analisados para obtenção de informações valiosas, que darão suporte à tomada de decisões estratégicas.

As informações obtidas a partir da extensa exploração e manipulação de dados se tornaram o bem mais precioso para as empresas. Ao longo dos últimos anos, para atender a demanda das áreas de negócios responsáveis pela tomada de decisão, e garantirem vantagem competitiva com a digitalização de processos de obtenção e análise de dados, as organizações foram forçadas a buscar por novas maneiras de alcançar as informações estratégicas (PONNIAH, 2010). Neste movimento, a tecnologia da informação ganhou destaque, fornecendo mecanismos de suporte a tomada de decisão, como *Data Mining*, *Business Analytics* e *Business Intelligence* (MISRA; RAHMAN, 2013), além de viabilizar o surgimento de ferramentas robustas para manipulação e visualização de dados, bem como da ciência de dados *etc.*

Considerados como o “coração” do *Data Warehouse*<sup>1</sup> - um repositório de dados centralizado que fornece suporte aos processos analíticos, os processos ETL cumprem uma função primordial na garantia da qualidade dos dados e na preparação dos mesmos para análise, sendo uma etapa essencial no ciclo de vida de um projeto de B.I (BISPO, 1998). Desta forma, é possível observar que a importância dos processos ETL se torna ainda mais evidente em um contexto de crescimento exponencial dos dados, os quais são provenientes das diversas fontes de dados

---

<sup>1</sup> *Data Warehouse* trata-se de um repositório de dados centralizado, projetado para dar suporte aos processos analíticos e às tomadas de decisões de uma organização. Para Kimball e Ross (2013), o DW se caracteriza como “um conjunto de dados orientado por assuntos, integrado, não volátil e variante no tempo, que dá suporte às decisões gerenciais”.



existentes, o que inclui sistemas ERP de fontes internas e externas, redes sociais, sensores e dispositivos *IoT* (*Internet of Things*, do português: Internet das Coisas).

Neste sentido, a usabilidade do T-SQL (*Transact-SQL*), uma variante da linguagem de consulta estruturada SQL (*Structured Query Language*), ganha destaque em possibilitar o desenvolvimento e manutenção dos processos ETL. O T-SQL é uma linguagem que interage com os bancos de dados *Microsoft SQL Server*, geralmente através dos Sistemas de Gerenciamento de Banco de Dados<sup>2</sup> (SGBD's) da *Microsoft SQL Server Management Studio* (SSMS) - que são sistemas gerenciamento responsáveis por oferecer recursos para a criação, organização, e manipulação de dados (SILBERSCHATZ; KORTH; SUDARSHAN., 2010). O ecossistema da *Microsoft* em questão (*SQL Server* e *SMSS*) é considerado uma das principais plataformas de armazenamento e processamento de dados em ambientes de *Business Intelligence* e *Business Analytics*, consolidando a 3ª posição no Ranking de SGBD's mais utilizado pelas organizações ao redor do mundo (DB-ENGINES RANKING, 2023).

Sob a perspectiva do *Business Intelligence* e *Business Analytics*, percebemos que os processos ETL desempenham um papel fundamental na garantia da qualidade dos dados, na integração de fontes diversas e na preparação dos mesmos para análise. Deste modo, através dos processos ETL, é possível obter um conjunto de dados consolidado, limpo e estruturado, permitindo a realização de análises mais precisas e confiáveis (KIMBAL; ROSS, 2013).

Além disso, os processos ETL oferecem uma base sólida para garantir a implementação de estratégias de B.I. Eles permitem que as diversas fontes de dados analíticos, como os *Data Warehouses*, *Data Marts*<sup>3</sup> e outras estruturas de armazenamento, sejam carregadas com dados

---

<sup>2</sup> **Sistemas de Gerenciamento de Bancos de Dados (SGBD's)** se caracterizam como sistemas de gerenciamento e controle de acesso aos dados que se encontram armazenados em um determinado banco de dados. São capazes de oferecer recursos para a criação, organização, e manipulação de dados e informações, garantindo a integridade e a segurança dos ambientes. Conforme Silberschatz, A., Korth, H. F.; Sudarshan, S. (2010), os SGBD's fornecem interfaces para a execução de consultas complexas, controle de transações e recuperação de falhas.

<sup>3</sup> **Data Marts** são subconjuntos de um *Data Warehouse*<sup>1</sup> que contém dados específicos para atender às necessidades de análise de um departamento, equipe ou projeto específico. Eles são projetados para fornecer acesso rápido e fácil aos dados relevantes, geralmente focando em um único assunto ou área de negócio. Segundo Kimball e Ross (2013), os *Data Marts* são estruturas dimensionais otimizadas para consultas analíticas, permitindo uma visão mais detalhada e específica dos dados em comparação com o DW. Eles são construídos usando técnicas de modelagem dimensional e são alimentados por meio de processos ETL (*Extração, Transformação e Carga*) a partir do DW central. Os *Data Marts* fornecem aos usuários finais um ambiente mais especializado e de fácil compreensão para análise de dados, facilitando a tomada de decisões (INMON, 2005).





---

tratados a partir de uma regra de negócio, proporcionando uma visão integrada e abrangente das informações organizacionais, além de ser um facilitador na transformação de dados brutos em informações significativas e relevantes (KIMBAL; ROSS, 2013). Estes processos envolvem etapas de limpeza, enriquecimento, padronização e agregação dos dados, tornando-os aptos para análise. Essa preparação adequada dos dados é fundamental para a obtenção de resultados confiáveis, bem como, *insights* valiosos para os analistas de dados e gestores.

É importante mencionar que, embora existam vários softwares específicos para ETL amplamente utilizados no mercado, como o *Pentaho Data Integration*, *SAS Data Integration Studio*, *IBM Information Server*, *Red Hat etc.*, além do *SSIS (SQL Server Integration Services)* - uma solução própria da *Microsoft* voltada a integração de um pacote ETL para *SQL Server* através do *SSMS* (TAVARES, 2013), este estudo dará ênfase ao uso do *T-SQL (Transact-SQL)* como uma linguagem de construção e manutenção dos processos ETL.

Ao destacarmos a importância dos processos ETL, é importante também mencionar a necessidade de uma abordagem eficiente e robusta para sua construção e manutenção, definindo uma estratégia e os requisitos essenciais (KIMBALL; ROSS, 2013). Neste contexto, a usabilidade da linguagem *T-SQL* nos processos ETL se torna relevante e acaba se consolidando como a ferramenta a ser explorada no presente estudo, uma vez que a linguagem em questão, naturalmente, possui recursos disponíveis que permitem ao profissional de dados realizar os mais diversos tipos de preparações de dados, adequando-os para atender as demandas necessárias. Em termos técnicos, valendo-se do *T-SQL* nos processos ETL, é possível aproveitar sua capacidade de lidar com consultas complexas, gatilhos, realizar operações de junção e filtragem, aplicar funções avançadas e assim, executar as tarefas de transformação de dados. Em resumo, isto contribui para a agilidade e flexibilidade na construção dos fluxos de dados, permitindo a adaptação às necessidades específicas e particulares de cada projeto de *Business Intelligence* e *Business Analytics*.

Deste modo, a ampla utilização do *Microsoft SQL Server* e *SMSS* como um sistema de gerenciamento de banco de dados confiável e escalável proporciona suporte adequado ao uso do *T-SQL* nos processos ETL, além de permitir que as organizações concentrem os processos de armazenamento e tratamento de dados em um único ambiente.

Em síntese, objetiva-se com este relato tecnológico apresentar a usabilidade da linguagem *T-SQL (Transact-SQL)* - uma variação da linguagem *SQL (Structured Query Language)*

---



---

desenvolvida para interagir com o banco de dados *Microsoft SQL Server* - na construção e manutenção de processos ETL (*Extract, Transform and Load*, do português: Extração, Transformação e Carga) nos ambientes de *Business Intelligence* e *Business Analytics* de uma empresa brasileira líder em um setor estratégico e, a partir disto, discutir acerca dos resultados obtidos e das limitações identificadas no processo.



---

## 2. REFERENCIAL TEÓRICO

Nesta seção, serão exploradas as diversas bases teóricas que versam e fundamentam o tema deste relato tecnológico, com o objetivo de fundamentar o estudo e tornar mais acessível a compreensão e o respaldo das conclusões. Serão abordados conceitos relacionados a *Business Intelligence e Business Analytics, ETL e Transact-SQL (T-SQL)*.

### 2.1. Business Intelligence e Business Analytics

A partir de um paralelo entre as definições estabelecidas por Davenport e Harris (2007) e Turban *et al.* (2008), os conceitos de *Business Intelligence (BI)* e *Business Analytics (BA)*, em geral, se caracterizam como processos essenciais para aquelas empresas que desejam aproveitar ao máximo os dados disponíveis – ou seja, os dados que são gerados pelas suas diversas fontes e posteriormente armazenados – para obter *insights* valiosos e tomar decisões pautadas em informações concretas e estratégicas. Ambas as abordagens têm o objetivo de reunir, avaliar e analisar os dados de um negócio, permitindo que as organizações monitorem seu desempenho por meio de KPI's<sup>4</sup>, identifiquem padrões, detectem tendências e antecipem oportunidades.

A importância de BI e BA reside no fato de que eles permitem que as empresas tomem decisões orientadas por informações extraídas a partir de dados, ao invés de depender exclusivamente de intuições ou suposições. Com base em informações precisas e relevantes, as organizações podem identificar áreas de melhoria, ajustar suas estratégias, otimizar processos e aumentar a eficiência operacional.

A seguir, exploraremos os conceitos de *Business Intelligence e Business Analytics*, bem como suas particularidades e relações.

---

<sup>4</sup> **Key Performance Indicators (KPIs)** são métricas que auxiliam os gestores na medição do desempenho da organização em relação a seus objetivos estratégicos. Eles são indicadores quantitativos que fornecem uma visão clara e mensurável do progresso e sucesso de uma empresa. Os KPIs são essenciais para o monitoramento e avaliação do desempenho, permitindo que as empresas identifiquem áreas de melhoria e tomem decisões informadas. (Fonte: Schwalbe, 2018).



### 2.1.1. Business Intelligence

*Business Intelligence* (B.I.) trata-se de um conceito que engloba diversas práticas, tecnologias e ferramentas voltadas para a coleta, organização, análise e apresentação de informações relevantes para a tomada de decisões empresariais. Diversos autores contribuíram para a definição e compreensão deste conceito em questão ao longo do tempo, dentre os quais, destacamos:

Richard M. Devens (DEVENS, 1868, p.210) foi um dos primeiros autores a utilizar o termo "*Business Intelligence*", referindo-se à capacidade de coletar informações sobre mercados, produtos e concorrentes para auxiliar na tomada de decisões para fins comerciais – obviamente, sem a utilização de recursos tecnológicos. Desde então, o conceito evoluiu e se tornou um estudo que abrange não apenas a coleta de informações, mas também a análise, interpretação e comunicação desses dados como um suporte essencial à gestão empresarial, valendo-se das ferramentas de tecnologia da informação como meio.

Foley e Guillemette (2010) destacam que o *Business Intelligence* envolve a transformação de dados brutos em informações úteis extraídas por meio de processos de coleta, armazenamento e análise. Essas informações podem ser apresentadas por meio de relatórios, painéis de controle ou *dashboards*, facilitando a visualização e interpretação dos dados.

Heinze (2014) destaca que o *Business Intelligence* se trata de uma abordagem estratégica para a gestão empresarial, permitindo que as organizações obtenham *insights* valiosos sobre seu desempenho, identifiquem tendências e padrões, e tomem decisões baseadas em evidências. A abordagem em questão envolve a integração de várias tecnologias e fontes de dados, como bancos de dados internos e externos, dados de mercado, dados de clientes e informações de mídia social, a fim de extrair conhecimento dos dados e obter uma visão abrangente do negócio.

Howson *et. al* (2019), define o *Business Intelligence* como um conjunto de metodologias, processos, arquiteturas e tecnologias que transformam dados brutos em informações úteis para a tomada de decisões empresariais. Rikhardsson e Yigitbasioglu (2018) enfatizam que o objetivo do B.I. é transformar dados em conhecimento, para então apoiar as organizações a entender e analisar seu desempenho, identificar oportunidades de mercado, antecipar tendências e melhorar sua eficiência operacional, além de claro, obterem vantagem competitiva.



---

Turban *et al.* (2008) apresenta uma definição mais abrangente de *Business Intelligence*, afirmando ser um conceito amplo que inclui arquiteturas, ferramentas, bancos de dados, aplicações e metodologias. Essa descrição ressalta a natureza abrangente do B.I. e sua aplicabilidade em diversos contextos empresariais.

Em resumo, o *Business Intelligence* é um processo que utiliza tecnologias, metodologias e processos para coletar, analisar e apresentar informações relevantes para a tomada de decisões empresariais. Esta abordagem permite que as organizações tenham uma visão mais clara de seu desempenho, identifiquem oportunidades de negócios e tomem decisões pautadas em evidências abstraídas por meio de informações.

No próximo tópico, iremos explorar o tema do *Business Analytics* e sua relação com o *Business Intelligence*.

### 2.1.2. Business Analytics

Partindo das definições gerais de Sharda *et al.* (2014) e Davenport e Harris (2007), podemos dizer que o *Business Analytics* é um conceito relacionado ao *Business Intelligence* que se concentra principalmente na aplicação de técnicas avançadas de análise de dados para extrair *insights* relevantes e acionáveis do conjunto de dados de uma organização. O B.A. vai além da simples coleta e análise de informações, buscando entender os padrões, as relações causais e as tendências ocultas nos dados para fornecer uma base concreta para a tomada de decisões estratégicas, valendo-se do extenso uso da tecnologia da informação e da competência dos analistas envolvidos - uma combinação eficaz que capacita o negócio a se tornar uma força competitiva dentro do seu setor (AKTER, 2016).

Alguns pontos que nos permitem compreender as principais diferenças entre os conceitos de *Business Intelligence* e *Business Analytics*, residem no alvo e na profundidade da análise. Enquanto o B.I se concentra principalmente em relatórios e análises descritivas do passado (FOLEY; GUILLEMETTE, 2010), o B.A. se estende além disso para análises preditivas e prescritivas, fornecendo *insights* que permitem impulsionar a estratégia e o planejamento futuros (CHOU, 2020). Em outras palavras, enquanto o B.I. é mais orientado para as operações diárias e o



monitoramento de desempenho, o B.A. tem um foco mais estratégico e de longo prazo, ajudando a orientar a inovação, bem como na tomada de decisões.

Para Chou (2020), além das tradicionais metodologias de coleta, análise de dados e disponibilização de painéis de visualização que envolvem o *Business Intelligence*, o *Business Analytics* alcança a capacidade de prever resultados futuros com base em dados históricos e tendências, graças ao uso de técnicas analíticas avançadas – e computacionais –, como os variados tipos de análise preditiva, *Machine Learning*<sup>5</sup> e Ciência de Dados. Tais capacidades analíticas possibilitam que as empresas antecipem eventuais mudanças no mercado, evasão de clientes, identifiquem riscos e oportunidades, como também, tomem medidas proativas, ao invés de reativas.

Davenport e Harris (2007) caracterizam o *Business Analytics* como "o extenso uso de dados, análises quantitativas, estatísticas descritivas e modelos preditivos para impulsionar a tomada de decisões e o desempenho empresarial". Esta definição destaca a importância do uso de técnicas analíticas avançadas, como modelagem estatística e mineração de dados, por exemplo, para obter uma compreensão mais profunda dos dados e obter *insights* significativos.

Lavalle *et al.* (2011) destacam que o *Business Analytics* tem como principal objetivo ajudar as organizações a tomar decisões informadas e baseadas em dados para impulsionar a inovação, melhorar a eficiência operacional e ganhar vantagens competitivas. Naturalmente, isto demanda o uso de ferramentas e tecnologias para analisar grandes volumes de dados estruturados e não estruturados, identificando padrões, correlações e *insights* com objetivo de orientar as organizações com eventuais tomadas de decisões.

Albright e Winston (2019) enfatizam que o *Business Analytics* está relacionado à capacidade de realizar análises preditivas e prescritivas, permitindo que as organizações tomem medidas antecipadamente com base em previsões futuras. Para isso, a importância da aplicação de algoritmos e técnicas avançadas de *Machine Learning* e Inteligência Artificial para identificação de padrões ocultos nos dados e previsão de resultados futuros.

Davenport (2006) destaca a importância do *Business Analytics* como sendo um processo que não se limita apenas à análise retrospectiva, mas também abrange a análise em tempo real e a

---

<sup>5</sup> *Machine Learning* trata-se de um subconjunto da inteligência artificial que envolve o desenvolvimento de sistemas os quais permitem que os computadores aprendam e melhorem seu conhecimento, e conseqüentemente, tomem decisões com base em dados, segundo Mitchell (1997).



---

análise preditiva. Desde modo, isto permite que as organizações tomem decisões mais rápidas e precisas, respondendo às eventuais mudanças no ambiente de negócios de forma ágil e antecipando assim, as tendências e oportunidades.

Por fim, Agrawal (2014) destaca que o *Business Analytics* é fundamental para a transformação digital das organizações, permitindo que elas se tornem mais orientadas por dados e baseadas em evidências. O autor enfatiza a importância de desenvolver uma cultura analítica dentro das organizações, onde os dados são valorizados, a análise é incorporada em todas as áreas de negócios e as decisões são baseadas em *insights* obtidos a partir da extensa investigação de dados.

## 2.2. ETL

Os processos ETL (*Extract, Transform and Load*, da tradução: Extração, Transformação e Carga) é essencial para a consolidação e preparação de dados nos ambientes que envolvam processos relacionados a *Business Intelligence* e *Business Analytics*. Segundo Kimball e Ross (2013), o ETL envolve a extração de dados de várias fontes, a transformação desses dados em um formato adequado para análise e por fim, a carga dos dados em um *Data Warehouse* ou em um *Data Mart*, por exemplo.

A etapa de extração se resume em coletar dados de diferentes fontes, como bancos de dados, arquivos *CSV*, *APIs*, entre outros. Sob a ótica de Tavares (2013), a etapa de extração consiste na coleta dos dados a partir de diferentes fontes, como bancos de dados transacionais, arquivos estruturados e não estruturados, sistemas legados etc. É importante salientar que os dados em questão podem estar armazenados em formatos distintos e, portanto, requerem técnicas específicas para sua extração e integração em um ambiente centralizado.

Os dados extraídos são então submetidos a um processo de transformação, que envolve limpeza, padronização, enriquecimento e agregação, visando garantir a qualidade e a integridade dos dados (SHARDA *et al.*, 2014). Esta etapa também pode, naturalmente, envolver a aplicação de regras de negócio e a criação de métricas personalizadas. Oliveira (2019) destaca a importância da limpeza dos dados para remover inconsistências, duplicações e valores inválidos. Além disso,



---

durante a transformação, os dados podem ser enriquecidos com informações adicionais, como funções para cálculos específicos, agregações, classificações etc.

Após a transformação, os dados são carregados no local de destino especificado, onde estarão à disposição para análise e tomada de decisões (SHARDA *et al.*, 2014). Tavares (2013) ressalta que este passo em questão, envolve a definição da estrutura do armazenamento, a criação de tabelas, índices e relacionamentos, além da aplicação de políticas de segurança e acesso aos dados. A carga pode ser realizada em tempo real ou em intervalos programados, dependendo dos requisitos de atualização dos dados e da ferramenta utilizada para os processos ETL.

Deste modo, os processos ETL desempenham um papel fundamental nos ambientes de BI e BA, pois garante a disponibilidade de dados confiáveis, consistentes e atualizados para análise e geração de *insights*, de acordo com SHARDA *et al.* (2014). Eles possibilitam, portanto, a integração de informações provenientes de diversas fontes e a criação de uma visão unificada dos dados, facilitando a análise de negócios e posteriormente, a tomada de decisões estratégicas (KIMBALL; ROSS, 2013).

É importante mencionar que os processos ETL enfrentam constantes desafios relacionados ao volume crescente de dados, à heterogeneidade das fontes, à complexidade das transformações e à necessidade de garantir a qualidade dos dados. Diversas ferramentas e técnicas foram desenvolvidas para auxiliar estes processos, como o uso do ETL automatizado, modelagem dimensional e a adoção de boas práticas de governança de dados (OLIVEIRA, 2019).

Neste sentido, é importante ressaltar que a escolha da ferramenta de ETL adequada varia conforme as necessidades específicas de cada projeto, das características das fontes de dados, como também, da mão de obra e orçamento disponíveis para tal. Como dito anteriormente, existem no mercado atualmente diversas ferramentas voltadas a criação, manutenção e automatização de processos ETL, no entanto, a ênfase deste relato tecnológico será na linguagem T-SQL, a qual será explorado no tópico seguinte.





---

### 2.3. *Transact-SQL* (T-SQL)

O *Transact-SQL* (T-SQL) se caracteriza como uma extensão da linguagem de consulta estruturada SQL (*Structured Query Language*) utilizada para interagir com os bancos de dados Microsoft *SQL Server*, segundo a Documentação Oficial da *Microsoft*.

Para Ben-Gan (2016), o *Transact-SQL* oferece uma ampla gama de recursos para consultas avançadas em bancos de dados, incluindo junções, consultas internas (ou subconsultas), agregações, ordenação e filtragem de dados. Ainda, possui recursos de manipulação de *strings*, datas e conversões de tipos de dados, bem como permite a criação de *stored procedures* (procedimentos armazenados), funções definidas pelo usuário e *triggers* (gatilhos) – ideais para os processos ETL apresentados neste relato tecnológico. Tais recursos facilitam a modularização do código e a reutilização de lógica de negócios, aumentando a eficiência do desenvolvimento e a manutenção do código, segundo a perspectiva de Grant Fritchey (2018).

Na visão de Korotkevitch (2016), a linguagem em questão oferece recursos para a administração de bancos de dados, como criação de tabelas, índices, visualização e gerenciamento de metadados, controle de transações e gerenciamento de segurança. Além disso, também fornece recursos para programação de tarefas agendadas, importação/exportação de dados, bem como a otimização de consultas para melhorar o desempenho do banco de dados.

### 3. SITUAÇÃO PROBLEMA

O presente relato tecnológico tem como objetivo apresentar a usabilidade da *Transact-SQL* como uma linguagem para desenvolvimento e manutenção dos processos ETL nos ambientes de *Business Intelligence* e *Business Analytics*. O cenário utilizado para o diagnóstico da situação problema a partir do tema principal, será o de uma empresa brasileira líder em um setor estratégico, responsável pela transformação de matéria-prima em produtos e sua comercialização para grandes clientes do mercado interno e externo.

Dada a natureza e porte da referida empresa, um volume massivo de dados é gerado diariamente a partir de diversas fontes, com destaque para os sistemas ERP que gerenciam módulos voltados para os diferentes setores da organização, como comercial, logística, recursos humanos, faturamento, plantas produtivas etc. Tais sistemas alimentam constantemente os servidores e bancos de dados da organização com informações importantes que, posteriormente, servirão como principal insumo para as ferramentas de análises de B.I. e B.A., e conseqüentemente, para geração de *insights* relevantes, apoiando as equipes e os gestores nas tomadas de decisão em diversos assuntos e temas que versam a realidade em que a organização se encontra inserida.

Neste sentido, a situação problema consiste nos desafios e limitações existentes para implementar e gerenciar um processo ETL em um fluxo de dados de forma eficiente, valendo-se da usabilidade da linguagem T-SQL e do Banco de Dados *SQL Server*, de modo a explorar os recursos do ecossistema em questão para realizar a extração, tratamento (padronização, limpeza e agrupamento) e carga dos dados em um *Data Warehouse*, sem utilizar ferramentas externas para construção e manutenção do processo ETL.

Para ilustrar o parágrafo anterior, abaixo encontra-se o fluxograma com as etapas envolvidas no contexto de fluxo de dados onde a situação problema reside:



Figura 1. Fluxograma do Fluxo de Dados em que a situação problema reside.

Fonte: Autor



---

#### 4. INTERVENÇÃO ADOTADA

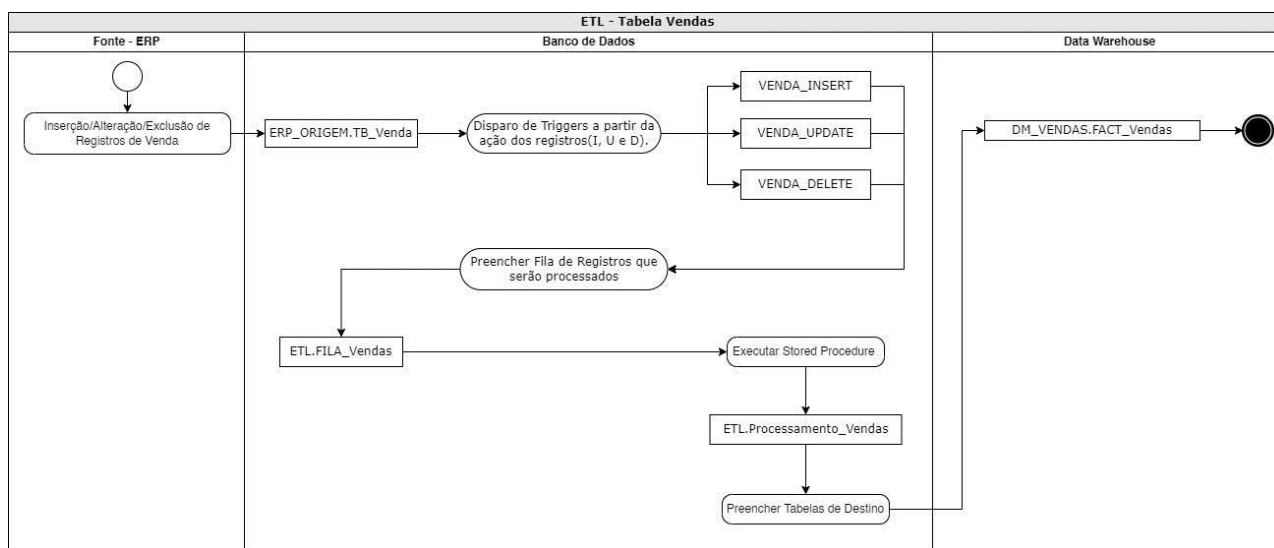
A partir da situação problema apresentada anteriormente, será adotada como proposta de intervenção a usabilidade da linguagem T-SQL na implementação de um processo ETL, perpassando pelas etapas de extração, tratamento e carga, para então disponibilizar os dados relativos a área comercial (vendas) oriundos de uma Tabela Origem Venda do Sistema ERP em uma Tabela Fato (Vendas) do *Data Warehouse*, dados estes que se encontrarão, posteriormente, à disposição das equipes responsáveis pelos ambientes de *Business Intelligence* e *Business Analytics* da organização.

##### 4.1. Objetivos

Para esclarecer o objetivo principal do relato tecnológico em questão, que reside em apresentar a usabilidade da *Transact-SQL* como uma linguagem para desenvolvimento e manutenção dos processos ETL nos ambientes de B.I. e B.A., bem como justificar a metodologia da intervenção adotada, faz-se necessário destacar dois pontos: **(i)** Que apesar da grande quantidade de ferramentas de ETL disponíveis no mercado e a vasta quantidade de recursos oferecidos por elas, a presente intervenção visa obedecer aos requisitos da situação problema, que se concentram no desenvolvimento de um processo através dos recursos que a organização dispõe no momento, ou seja, o *SQL Server*, *SMSS* e conseqüentemente, a T-SQL; **(ii)** Que existem no cenário atual da organização muitas de tabelas de dados oriundas de diversas fontes distintas, com características específicas, cujos dados são submetidos aos processos ETL por meio da linguagem T-SQL. Desta forma, cada processo é desenvolvido de uma maneira distinta, de modo a atender uma determinada particularidade na preparação dos dados para sua área demandante. Assim, será apresentado por meio área de vendas a maneira como a linguagem T-SQL é utilizada para desenvolver e gerenciar um processo ETL nos ambientes de B.A e B.I, demonstrando o planejamento e desenvolvimento de suas etapas, destacando também, os principais recursos da linguagem em questão que são utilizados para possibilitar a extração, transformação, e por fim, a consolidação dos dados em um DW.

## 4.2. Metodologia

Para melhor compreensão do fluxograma apresentado na Figura 1, bem como do processo ETL dos dados da Tabela Origem Venda para a Tabela Fato (Vendas) que ocorrerá especificamente nos ambientes de “Banco de Dados” e “Data Warehouse”, ilustramos na Figura 2, abaixo, o diagrama BPM (*Business Process Modeling Notation*, da tradução: Notação de Modelagem de Processos de Negócios), apresentando uma visão geral a partir das etapas, objetos e ações contidas no fluxo de dados onde o processo ETL será desenvolvido. Posteriormente, faremos o detalhamento de cada etapa e tabela envolvida no BPM abaixo, a fim de contextualizar a metodologia abordada:



**Figura 2.** BPM – Visão detalhada do fluxo de dados e do processo ETL.

Fonte: Autor

Através da Figura 3, o Modelo Entidade Relacionamento (MER) fornece uma visão ampla das demais tabelas presentes no banco de dados de origem, que através de seus atributos de identificação, se relacionam à Tabela Origem Venda “ERP\_ORIGEM\_TB.Venda”. Esta representação é importante para compreendermos a estrutura da Tabela Origem Venda, que será submetida ao processo ETL, bem como seu relacionamento com as demais tabelas do banco de dados de origem.

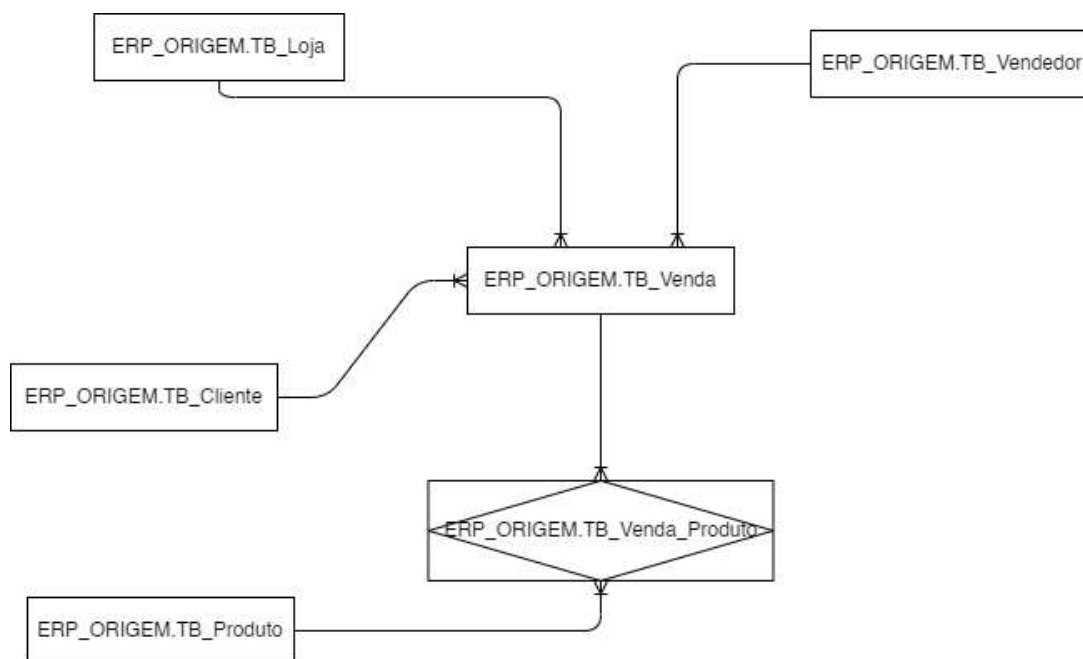


Figura 3. MER – (Banco de Dados de Origem) Tabela Origem Venda e entidades relacionadas.  
Fonte: Autor

Ao analisar o MER da Figura 3, podemos identificar as entidades envolvidas (tabelas), seus relacionamentos e as chaves que as vinculam à Tabela Origem Venda, permitindo uma compreensão abrangente do fluxo de dados da origem e das interações entre as diferentes tabelas provenientes do banco de dados do Sistema ERP. Esta compreensão é fundamental para garantir a integridade e o planejamento da transformação dos dados durante o processo de ETL a partir da linguagem T-SQL. Abaixo, a descrição dos campos e tipos de dados que compõem Tabela Origem Venda “ERP\_ORIGEM\_TB.Venda”:

Campo	Tipo	Descrição
pk_venda	int	Chave primária gerada para identificar exclusivamente cada venda.
data_venda	varchar	Data em que a venda foi realizada.
fk_vendedor	int	Chave estrangeira que se relaciona com a tabela de vendedores, identificando o “id” do vendedor associado à venda.
fk_cliente	int	Chave estrangeira que se relaciona com a tabela de clientes, identificando o “id” do cliente associado à venda.
fk_loja	int	Chave estrangeira que se relaciona com a tabela de lojas, identificando o “id” da loja onde a venda ocorreu.

Apesar dos conceitos e métodos concernentes à modelagem de tabelas *do Data Warehouse* não serem o foco principal do presente relato tecnológico, apresentamos na Figura 4, como forma de complementação ao tema e às Figuras 2 e 3, o DER (Diagrama Entidade Relacionamento) que compreende a estrutura das Tabela Fato (Vendas) que será carregada no DW a partir do ETL desenvolvido pela linguagem T-SQL. Além disso, o DER também ilustra as Tabelas Dimensões, que apesar de não estarem diretamente envolvidas no processo ETL apresentado, se relacionam com as Tabela Fato do DW em questão.

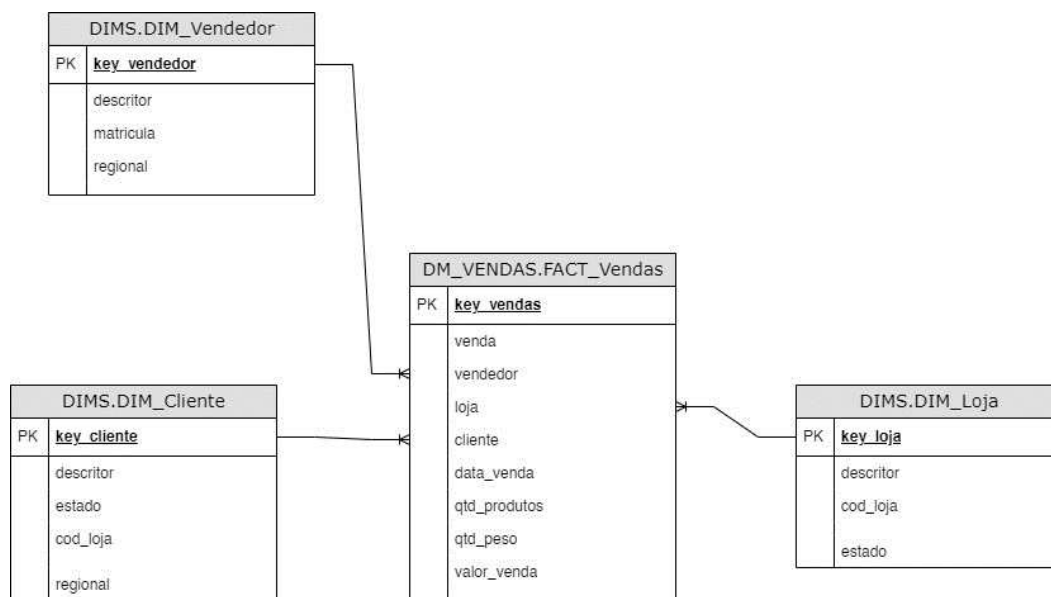


Figura 4. DER – (DW) Tabela Fato e suas dimensões.  
Fonte: Autor.

O DER em questão (Figura 4), representa a estrutura e as relações entre as Tabela Fato e suas Dimensões a partir de uma modelagem estrela, permitindo que sejam realizadas análises e consultas dos dados de vendas considerando diferentes dimensões, como vendedor, cliente e loja. Abaixo, a descrição dos campos e tipos de dados das tabelas apresentadas no DER em questão:



### I. Tabela Fato "DM\_VENDAS.FACT\_Vendas":

Campo	Tipo	Descrição
key_vendas	int	Chave primária para identificar cada venda.
venda	int	Número de identificação da venda.
vendedor	int	Identificação do vendedor responsável.
loja	int	Identificação da loja onde a venda ocorreu.
cliente	int	Identificação do cliente associado à venda.
data_venda	datetime	Data em que a venda foi realizada.
qtd_produtos	int	Quantidade de produtos
qtd_peso	decimal	Peso dos produtos vendidos na transação.
valor_venda	decimal	Valor total da venda.

### II. Tabela Dimensão "DIMS.DIM\_Vendedor":

Campo	Tipo	Descrição
key_vendedor	int	Chave primária para identificar cada vendedor.
descriptor	varchar	Descrição do vendedor.
matricula	varchar	Matrícula ou identificação única do vendedor.
regional	varchar	Regional à qual o vendedor está atribuído.

### III. Tabela Dimensão "DIMS.DIM\_Cliente":

Campo	Tipo	Descrição
key_cliente	int	Chave primária para identificar cada cliente.
cod_cliente	varchar	Código para identificar o cliente que realizou a compra.
descriptor	varchar	Descrição do cliente.
estado	varchar	Estado onde o cliente está localizado.
pais	varchar	País onde o cliente está localizado.

### IV. Tabela Dimensão "DIMS.DIM\_Loja":

Campo	Tipo	Descrição
key_loja	int	Chave primária para identificar cada loja.
cod_loja	varchar	Código para identificar a loja onde ocorreu a venda.
descriptor	varchar	Descrição da loja.
regional	varchar	Regional à qual a loja está atribuída.

A partir das definições e conceitos acerca da Tabela Origem Venda, Tabela Fato (Vendas) e Dimensões (Loja, Cliente e Vendedor) relacionadas, bem como da visão geral ilustrada na Figura 2, passamos então a definir as etapas, objetos e ações que perpassam pelos ambientes descritos no fluxo de dados, como a Fonte de Dados – Sistema ERP, Banco de Dados *SQL Server* e *Data Warehouse*, que também pertence à instancia *SQL Server*.



	Ambiente	Etapa	Descrição
1	Fonte de Dados – Sistema ERP	Inserção, Atualização ou Deleção de Dados de Vendas	A partir da interação com a interface de módulo comercial do Sistema ERP, o usuário registra informações que envolvem, em termos de banco de dados, operações como Inserção (I – <i>Insert</i> ), Atualização (U – <i>Update</i> ) ou Deleção (D – <i>Delete</i> ) de dados de venda. Estes dados são então carregados juntamente com seus dados, em uma tabela do banco de dados, destinada ao armazenamento de informações oriundas das transações/operações do módulo em questão, de acordo com os seus operações (I, U ou D).
2	Banco de Dados ( <i>SQL Server</i> )	Dados oriundos do Sistema ERP são Inseridos, Atualizados ou Deletados da Tabela Origem Venda	Os dados são então Inseridos, Atualizados ou Deletados na Tabela Origem Venda do Banco de Dados, através das operações (U, I e D) oriundos do Sistema ERP.
3	Banco de Dados ( <i>SQL Server</i> )	Disparo de <i>Triggers</i> <sup>6</sup> (Gatilhos) conforme operação dos dados (I, U e D).	Nesta etapa, as <i>Triggers</i> ( <i>VENDA_INSERT</i> , <i>VENDA_UPDATE</i> e <i>VENDA_DELETE</i> ) relacionadas a Tabela Origem Venda são configuradas para serem disparadas a partir de uma ação específica, que no caso, é a operação (I, U ou D) o qual os dados sofreram na Tabela Origem Venda. O objetivo é capturar a chave “pk_venda” dos dados, juntamente com a operação que sofreram na Tabela Origem Venda, para então carregá-los na Tabela Fila “ETL.FILA Vendas”.
4	Banco de Dados ( <i>SQL Server</i> )	Chaves e operações dos dados entram na tabela de fila de processamento “ETL.FILA_Vendas”.	As chaves e operações dos dados oriundos da Tabela Origem Venda são carregados na Tabela Fila para entrarem em uma fila de processamento. A tabela em questão é desenvolvida a partir das definições de TAD – Tipos Abstratos de Dados, que obedece ao conceito FIFO ( <i>First In, First Out</i> ), baseado em agrupamento por ações, ou seja, os primeiros “dados” a entrarem na tabela fila são os primeiros a saírem para processamento, agrupados pelas ações D, para deleção, U para atualização e I para Inserção. É importante ressaltar que os dados brutos oriundos da Tabela Origem Venda não são armazenados na Tabela Fila, apenas suas chaves e operações, como dito anteriormente. Isto serve para indicar a SP ( <i>Stored Procedure</i> ), na etapa posterior, quais são os dados que deverão ser extraídos, tratados e carregados no DW. Para gerenciamento dos dados que se encontram na Tabela Fila, existem colunas específicas, tais como: <b>Coluna “Operação”:</b> destinada a identificação do tipo de operação que os dados sofreram para serem capturados pelas <i>Triggers</i> na Etapa 3. A coluna indicará (I), para os dados que sofreram Inserção, (U) para os que sofreram Atualização e (D) para aqueles que foram deletados na Tabela Origem Venda; <b>Coluna “Entrada Fila”:</b> destinada a informação referente a data e horário em que dado entrou na Tabela Fila; <b>Coluna “Status”:</b> destinada a indicar a situação do processamento dos dados, onde os que estão aguardando processamento são classificados como “Em Fila”, os que estão em

<sup>6</sup> *Triggers* (da tradução: gatilhos) se trata de uma estrutura do banco de dados que opera como uma função, sendo disparada mediante alguma ação. Geralmente essas ações que disparam os *triggers* são alterações nas tabelas por meio de operações de inserção, deleção e atualização de dados (I - *Insert*, D - *Delete* e U - *Update*). Fonte: DevMedia.





			processamento como “Processando”, os que foram processados com sucesso “Processado” e aqueles que porventura apresentaram algum erro durante o processamento como “Erro”.
5	Banco de Dados (SQL Server)	<ul style="list-style-type: none"><li>i. Execução do <i>Job SQL Server Agent</i><sup>7</sup></li><li>ii. Execução de <i>Stored Procedure</i><sup>8</sup></li><li>iii. Preenchimento das Tabela Fato no <i>Data Warehouse</i></li></ul>	<p>Enquanto as etapas anteriores ocorrem na medida em que os dados são Inseridos, Atualizados ou Deletados na Tabela Origem a partir do Sistema ERP e carregados na Tabela Fila para o processamento, nesta etapa, a execução de um <i>Job</i> de ETL Comercial do <i>SQL Server</i> (i) é configurada para obedecer a uma periodicidade pré-determinada.</p> <p>O <i>Job</i> tem a função de executar uma <i>Stored Procedure</i> (Procedimento Armazenado) “ETL.Processamento_Vendas” (ii) a cada 02 minutos, a qual concentra um <i>script</i> para execução de uma série de instruções/tarefas.</p> <p>A partir das chaves e operações (U, I e D) dos dados que se encontram na Tabela Fila, a SP acessa a Tabela Origem Venda e os extrai, juntamente com seus dados, para a partir das instruções codificadas em T-SQL, os mesmos possam ser submetidos ao tratamento, limpeza, padronização e agrupamento com as Tabelas Dimensões do DW.</p> <p>Logo após os dados e seus dados serem submetidos às instruções de extração e tratamento (limpeza, padronização e agrupamento) contidas na SP, são então carregados (iii) para sua tabela de destino correspondente.</p> <p><b>Portanto, nesta etapa (5) ocorre efetivamente a Extração, Transformação e a Carga do processo ETL.</b></p>
6	<i>Data Warehouse</i> (SQL Server)	Dados disponibilizados nas Tabela Fato do DW	Neste ponto, os dados submetidos ao processo ETL, encontram-se devidamente tratados, padronizados, limpos e agrupados em suas tabelas de destinos, aptos para serem utilizados pelas ferramentas analíticas.

### 4.3.Desenvolvimento

No desenvolvimento e manutenção do processo ETL da Tabela Venda e seus respectivos dados, será utilizado o Sistema de Gerenciamento de Banco de Dados (SBGD) *Microsoft SQL Server Management Studio (Versão 15.0)*, e do banco de dados *Microsoft SQL Server 2016 Enterprise Edition (SP3)*, (*Versão 13.0.6*), utilizando como linguagem de interação ao banco de

<sup>7</sup> *SQL Server Agent* é o responsável pela automação do *SQL Server*, trabalhando com o processamento de tarefas previamente agendadas, como os *Jobs*, que se caracterizam como uma série especificada de operações executadas sequencialmente, como funções e scripts do banco de dados. Fonte: Documentação Oficial da Microsoft.

<sup>8</sup> *Stored Procedure* trata-se de um grupo de uma ou mais instruções do *Transact-SQL*, que podem ser executadas como um único lote de código, a partir de uma única chamada. As SP podem conter instruções de programação que executam operações no banco de dados, o que inclui a chamada de outros procedimentos e o retorno de um valor de status a um programa de chamada para indicar êxito ou falha (e o motivo da falha). Fonte: Documentação Oficial da Microsoft



dados a T-SQL, bem como os recursos disponíveis pela linguagem em questão e as boas práticas de programação.

Com efeito, as figuras abaixo ilustram a concretização e o desenvolvimento das etapas descritas no item 4.2.:

	pk_venda	dat_venda	fk_cliente	fk_vendedor	fk_loja
1	90756	20150105 163322000	15033	200	LU11
2	90761	20150116 144001000	15037	900	LU11
3	90777	20150220 090051000	15042	510	LC01
4	90779	20150220 090140000	15042	510	LC01
5	90855	20150304 095439000	14677	210	LU08
6	90857	20150304 110822000	15043	510	LU11
7	90859	20150304 111043000	15043	750	LU11
8	90969	20150615 163401000	15028	200	LU11
9	91115	20150828 145635000	15125	510	LU11
10	91301	20151019 122606000	15140	200	LU11
11	91305	20151019 122718000	15140	200	LU11
12	91333	20151020 111812000	15142	200	LU11
13	91337	20151020 112225000	15142	200	LU11
14	91383	20151027 133857000	15146	200	LU11

Figura 5. Tabela Origem Venda “ERP\_ORIGEM\_TB.Venda” com seus campos e dados.

Fonte: Autor.

	pk_venda	des_prod	peso_prod	vlr_peso	qt_prod
1	90756	*****	11.000	24517.31	2
2	90761	*****	11.000	24517.31	5
3	90777	*****	11.000	24517.31	5
4	90779	*****	11.000	24517.31	8
5	90855	*****	11.000	24517.31	8
6	90857	*****	11.000	24517.31	8
7	90859	*****	11.000	24517.31	9
8	90969	*****	11.000	24517.31	6
9	91115	*****	11.000	24517.31	8
10	91301	*****	26.000	68775.45	7
11	91305	*****	26.000	68775.45	1
12	91333	*****	26.000	68775.45	2
13	91337	*****	26.000	68775.45	31
14	91383	*****	11.000	24517.31	4
15	91387	*****	11.000	24517.31	9

Figura 6. Tabela Origem Venda-Produto “ERP\_ORIGEM\_TB.Venda\_Produto” relacionada a Tabela Origem Venda com seus campos e dados.

Fonte: Autor.

Através da Figura 6, acima, ilustramos a Tabela Origem Venda-Produto “ERP\_ORIGEM\_TB.Venda\_Produto”. A Tabela em questão foi inicialmente apresentada no MER da Figura 3, como uma entidade relacionada a Tabela Origem Venda “ERP\_ORIGEM\_TB.Venda” (Figura 5), que será submetida ao processo ETL. A Tabela Origem Venda-Produto tem o propósito de relacionar as vendas com os produtos vendidos, além de apresentar o valor unitário do produto, seu peso e quantidade vendida. Assim, a partir da chave “pk\_venda”, a tabela em questão fornecerá dados complementares ao processo ETL, permitindo que a *Stored Procedure* “ETL.Processamento\_Vendas” possa, inclusive, calcular o valor total da venda e inserir na coluna da correspondente da Tabela Fato "DM\_VENDAS.FACT\_Vendas".

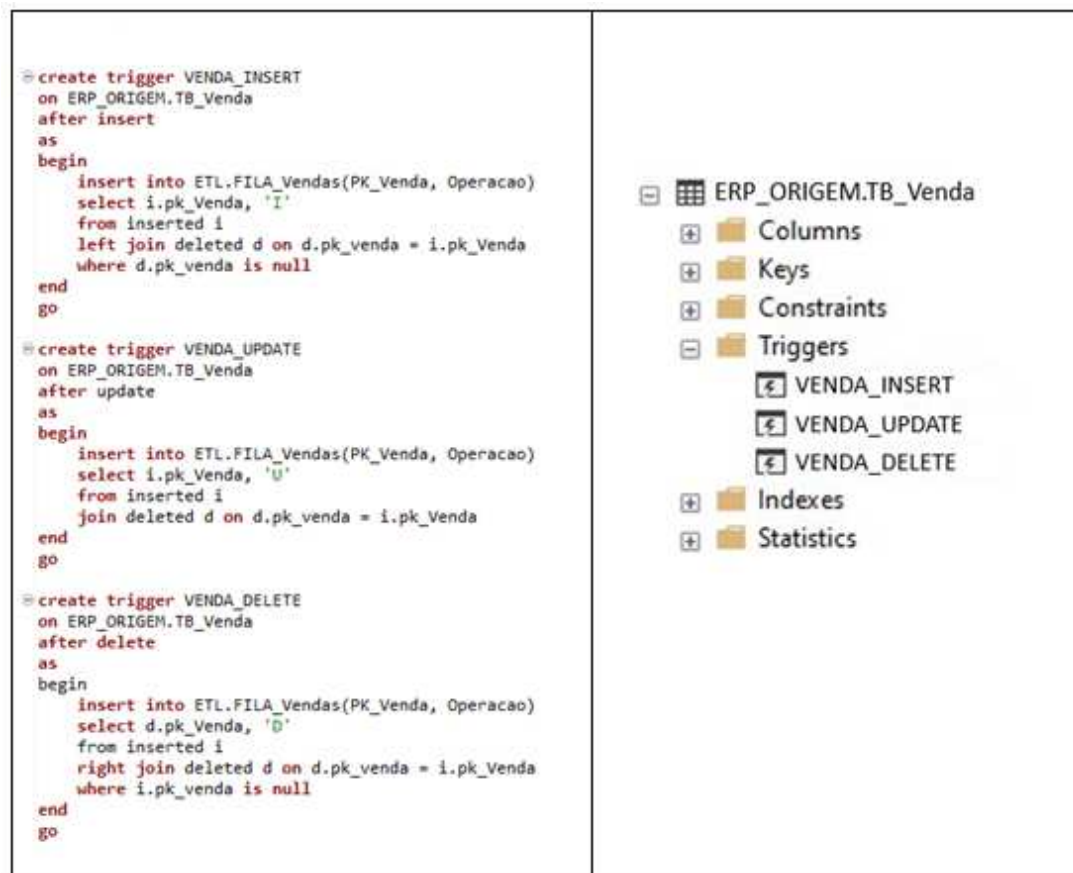


Figura 7. Script de criação das Triggers I, U e D vinculadas a Tb. Origem Venda (Esquerda) e Object Explorer do SSMS ilustrando as Triggers vinculadas a tabela em questão, após criadas. (Direita).

Fonte: Autor.



Na figura 7, acima, ilustramos as *Triggers* (VENDAS\_I, VENDAS\_U e VENDAS\_D) responsáveis por capturar as chaves “pk\_venda” dos dados da Tabela Origem Venda "ERP\_ORIGEM.TB\_Venda" e lança-las na Tabela Fila “ETL.FILA\_Vendas” (Figura 8) com suas respectivas operações (U, I e D) para processamento.

	PK_VENDAS	Operacao	Entrada_fila	Status
1	90756	I	2015-01-05 16:33:27	Processado
2	90761	I	2015-01-16 14:40:06	Processado
3	90777	I	2015-02-20 09:00:56	Processado
4	90779	I	2015-02-20 09:01:45	Processado
5	90855	I	2015-03-04 09:54:44	Processado
6	90857	I	2015-03-04 11:08:27	Processado
7	90859	I	2015-03-04 11:10:48	Processado
8	90969	I	2015-06-15 16:34:06	Processado
9	91115	I	2015-08-28 14:56:40	Processado
10	91301	I	2015-10-19 12:26:11	Processado
11	91305	I	2015-10-19 12:27:23	Processado
12	91333	I	2015-10-20 11:18:17	Processado
13	91337	I	2015-10-20 11:22:30	Processado
14	91383	I	2015-10-27 13:39:02	Processado
15	91387	I	2015-10-27 14:22:14	Processado

*Figura 8. Tabela Fila "ETL.FILA\_Vendas" ilustrando as chaves dos dados, suas operações, data e hora de entrada na fila e status.*

*Fonte: Autor.*

Na Figura 9, abaixo, apresentamos um trecho da operação (I) *Insert* realizada pela *Stored Procedure* "ETL.Processamento\_Vendas", ilustrando através da linguagem T-SQL como os comandos responsáveis pela extração, tratamento e carga na Tabela Fato Venda "DM\_VENDAS.FACT\_Vendas" são codificados. Além disso, é possível verificar comandos de agrupamento através das tabelas complementares, como a Tabela Origem Venda-Produto "ERP\_ORIGEM\_TB.Venda\_Produto" e as Tabelas Dimensões "DIMS.DIM\_Vendedor", "DIMS.DIM\_Cliente" e "DIMS.DIM\_Loja" para preencher os campos da Tabela Fato com seus dados correspondentes.



```
UPDATE b
SET b.status = 'Processando'
FROM ****.etl.fila_venda b
WHERE b.operacao = 'I'
AND b.status = 'Em Fila'

BEGIN TRY

INSERT INTO ****.dm_vendas.fact_vendas (venda, data_venda, vendedor, cliente, qtd_produtos, qtd_peso, valor_venda, loja)
OUTPUT inserted.venda, 'I' INTO #processamento
SELECT fvp.*
FROM ****.etl.fila_venda b
CROSS APPLY (
SELECT v.venda
, CONVERT(datetime, SUBSTRING(v.data_Venda, 1, 4) + '-' + SUBSTRING(v.data_Venda, 5, 2) + '-' + SUBSTRING(v.data_Venda, 7, 2) + ' ' +
SUBSTRING(v.data_Venda, 9, 2) + ':' + SUBSTRING(v.data_Venda, 11, 2) + ':' + SUBSTRING(v.data_Venda, 13, 2) + '.' +
SUBSTRING(v.data_Venda, 15, 3)) AS data_venda
, vd.key_vendedor
, c.key_cliente
, COUNT(vp.fk_produto) qtd_prod
, SUM(vp.peso) qtd_peso
, SUM(p.valorUnitario * ISNULL(vp.peso, COUNT(vp.fk_produto) OVER(PARTITION BY vp.fk_venda))) valor
, lj.key_loja
FROM ****.erp_origem.tb_venda v
JOIN ****.erp_origem.tb_venda_produtos vp ON v.pk_venda = vp.fk_venda
JOIN ****.erp_origem.tb_produtos p ON p.pk_produto = vp.fk_produto
JOIN ****.dims.dim_cliente c ON c.cod_cliente = v.fk_cliente
JOIN ****.dims.dim_vendedor vd ON vd.matricula = v.fk_vendedor
JOIN ****.dims.dim_loja lj ON lj.cod_loja = v.fk_loja
WHERE vp.venda = v.venda
) fvp
WHERE b.operacao = 'I'
AND b.status = 'Processando'

UPDATE b
SET b.status = 'Processado'
FROM ****.etl.fila_venda b
JOIN #processamento d ON b.pk_Venda = d.acao = b.operacao
WHERE b.operacao = 'I'

END TRY
BEGIN CATCH

UPDATE b
SET b.status = 'Erro'
FROM ****.etl.fila_venda b
WHERE b.operacao = 'I'
AND b.status = 'Processando'

END CATCH
```

Figura 9. Trecho da operação (I) Insert realizada pela Stored Procedure "ETL.Processamento\_Vendas".  
Fonte: Autor.

Por fim, na Figura 10, é possível observar a Tabela Fato "DM\_VENDAS.FACT\_Vendas" com seus respectivos campos e dados oriundos do processo ETL, devidamente tratados e agrupados.

	key_vendas	venda	data_venda	vendedor	loja	cliente	qtd_produtos	qtd_peso	valor_venda
1	90756	90756	2015-01-05 16:33:22.000	1	3	3	2	22.000	4457,7
2	90761	90761	2015-01-16 14:40:01.000	6	3	4	5	55.000	11144,25
3	90777	90777	2015-02-20 09:00:51.000	4	1	5	5	55.000	11144,25
4	90779	90779	2015-02-20 09:01:40.000	4	1	5	8	88.000	17830,8
5	90855	90855	2015-03-04 09:54:39.000	2	2	1	8	88.000	17830,8
6	90857	90857	2015-03-04 11:08:22.000	4	3	6	8	88.000	17830,8
7	90859	90859	2015-03-04 11:10:43.000	5	3	6	9	99.000	20059,65
8	90969	90969	2015-06-15 16:34:01.000	1	3	2	6	66.000	13373,1
9	91115	91115	2015-08-28 14:56:35.000	4	3	7	8	88.000	17830,8
10	91301	91301	2015-10-19 12:26:06.000	1	3	8	7	182.000	18516,47
11	91305	91305	2015-10-19 12:27:18.000	1	3	8	1	26.000	2645,21
12	91333	91333	2015-10-20 11:18:12.000	1	3	9	2	52.000	5290,42
13	91337	91337	2015-10-20 11:22:25.000	1	3	9	31	806.000	82001,51
14	91383	91383	2015-10-27 13:38:57.000	1	3	10	4	44.000	8915,4
15	91387	91387	2015-10-27 14:22:09.000	1	3	11	9	99.000	20059,65

Figura 10. Tabela Fato "DM\_VENDAS.FACT\_Vendas" com seus campos e dados após processo ETL.  
Fonte: Autor.



---

## 5. RESULTADOS OBTIDOS

A usabilidade da linguagem T-SQL na construção e manutenção dos processos ETL demonstrou que, de maneira geral, a obtenção de dados relevantes de várias fontes e a transformação dos mesmos em formatos adequados para análise é possível a partir da intensa exploração dos recursos que a linguagem e seu ecossistema oferecem.

Um dos principais pontos observados é que todo processo se concentrou diretamente nos ambientes do *SQL Server* e *SSMS*, sem a necessidade de um SGBD ou ferramenta ETL adicional. Assim, não houve custos referentes a aquisição de ferramentas específicas para construção e manutenção dos processos ETL, bem como para adaptação de infraestrutura e capacitação de pessoal etc. Ainda, é importante destacar os ganhos relativos ao tempo de planejamento, integração e execução, uma vez que as equipes envolvidas nos processos ETL, naturalmente, já estão familiarizadas com as estruturas, objetos e operações dos referidos ambientes.

Todos os recursos disponíveis pela linguagem, banco de dados e SGBD puderam se juntar e viabilizar a disponibilização de dados tratados sobre vendas nas Tabela Fato e Dimensões do *Data Warehouse*, permitindo com que os responsáveis pelos ambientes de B.A. e B.I. possam acessar esses dados de forma eficiente e realizar análises através das ferramentas analíticas. Deste modo, os dados submetidos ao processo ETL através da linguagem T-SQL em questão, poderão posteriormente, fornecer informações relevantes sobre as vendas, permitindo identificar tendências, padrões de consumo, segmentar clientes, avaliar desempenho de produtos e compreender o impacto de diferentes estratégias de negócio. Portanto, temos uma evidência de que a usabilidade da linguagem T-SQL em uma das etapas do fluxo de dados, possibilitou fortalecer a capacidade da organização em tomar decisões baseadas em evidências e *insights*.

Contudo, os desafios também estiveram presentes nesta implementação, e permitiram identificar alguns gargalos no processo em questão, como por exemplo: **(i)** a complexidade em codificar todo processo ETL de maneira a garantir a integridade de desempenho e fluidez na extração, tratamento e carga dos dados; **(ii)** no gerenciamento de erros e exceções; e por fim, **(iii)** na escalabilidade dos processos, ou seja, na medida em que a quantidade de dados aumenta e os requisitos de negócio evoluem a partir de novas demandas, o processo como um todo, se torna mais complexo e desafiador ao analista e desenvolvedor.



---

## 6. CONSIDERAÇÃO FINAIS

Ao explorar os recursos da linguagem T-SQL para o desenvolvimento e manutenção de processos ETL, bem como de seu ecossistema *Microsoft*, que inclui o *SQL Server* e o *SQL Server Management Studio* (SSMS), foi possível observar que os dados puderam ser tratados e disponibilizados com sucesso nas tabelas do *Data Warehouse* da organização, sem a necessidade de utilizar um *software* específico para tal, o que demandaria custos adicionais referentes a aquisição da ferramenta escolhida, adaptação de infraestrutura e capacitação de pessoal.

Apesar dos desafios e limitações impostas na utilização da T-SQL e de seu ecossistema, o processo em si, não somente aplicado à vendas, como no caso relatado, mas também nas demais áreas, permite com que as equipes de *Business Intelligence* e *Business* se beneficiem de dados padronizados, limpos e agrupados e realizem assim, análises avançadas para posteriormente, tomarem decisões embasadas em tendências, segmentações, desempenho etc.

Deste modo, a usabilidade da T-SQL se mostrou viável e promissora para o atual cenário da organização onde o estudo prático foi conduzido, como também para as demais empresas que porventura possuam o ecossistema *Microsoft*, bem como analistas e desenvolvedores familiarizados com a linguagem em questão a sua disposição. Deste modo, os recursos oferecidos pela linguagem e seu ecossistema colaboram para que as organizações se tornem orientadas a dados. Com conhecimento especializado, boas práticas, habilidades analíticas e uma abordagem adaptável, é possível superar os desafios impostos e obter os benefícios da T-SQL nos processos ETL.



---

## REFERÊNCIAS BIBLIOGRÁFICAS

- AGRAWAL, D. **Analytics Based Decision Making**. Journal of Indian Business Research, 2014.
- ALBRIGHT, S. C.; WINSTON, W. L. **Business Analytics: Data Analysis & Decision Making**. Cengage Learning, 2019.
- BEN-GAN, Itzik. **T-SQL Fundamentals**. 3ª edição. Microsoft Press, 2016.
- BISPO, C. A. F. **Uma Análise da Nova Geração de Sistemas de Apoio à Decisão**. Escola de Engenharia de São Carlos da Universidade de São Paulo, São Carlos, 1998.
- CHOU, L. **Top 10 Key Features of BI Tools in 2020**. Medium, 2020.
- DAMA INTERNATIONAL. **DAMA-DMBOK: Data Management Body of Knowledge**. 2nd Edition. Technics Publications, 2017.
- DAVENPORT, T. H.; HARRIS, J. G. **Competing On Analytics: The New Science Of Winning**. Boston, MA, EUA. Harvard Business Review Press, 2007.
- DAVENPORT, T.H. **Competing on Analytics**. Harvard Business Review, 2006.
- DB-ENGINES. **Ranking of Database Management Systems**. DB-Engines, Áustria, 2023. Disponível em: <https://db-engines.com/en/ranking>. Acesso em: 22 maio 2023.
- DEVENS, R. M. **Cyclopaedia of Commercial and Business Anecdotes: Comprising Interesting Reminiscences and Facts, Remarkable Traits and Humors... of Merchants, Traders, Bankers... Etc. in All Ages and countries...** Google Books. D. Appleton and Company, 1868.
- DEVMEDIA. **Triggers no SQL Server: Teoria e Prática Aplicada em uma situação real**. DevMedia, 2013. Disponível em: <https://www.devmedia.com.br/triggers-no-sql-server-teoria-e-pratica-aplicada-em-uma-situacao-real/28194>
- FOLEY, E.; GUILLEMETTE, M. G. **What is Business Intelligence?** International Journal of Business Intelligence Research, 2010.
- FRITCHEY, Grant. **SQL Server Execution Plans**. 3ª edição. Red Gate Books, 2018.
- HEINZE, J. **History of Business Intelligence**. BI Software Insight, 2014.
- HOWSON, C., RICHARDSON, J., SALLAM, R.; KRONZ, A. **Gartner Magic Quadrant for Analytics and Business Intelligence Platforms**. Gartner, 2019.





---

INMON, W. H. **Building the Data Warehouse**. John Wiley & Sons, 2005.

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

KIMBALL, R.; ROSS, M. **The Data Warehouse Toolkit: The Definitive Guide To Dimensional Modeling**. John Wiley & Sons, 2013.

KLEPPMANN, M. **Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, And Maintainable Systems**. O'Reilly Media, 2017.

KOROTKEVITCH, Dmitri. **Pro SQL Server Internals**. 3ª edição. Apress, 2016.

LAVALLE, S., LESSER, E., SHOCKLEY, R., HOPKINS, M., S.; KRUSCHWITZ, N. **Big Data, Analytics and the Path from Insights to Value**. MIT Sloan Management Review, 2011.

MARZ, N.; WARREN, J. **Big data: Principles and Best Practices of Scalable Real-Time Data Systems**. Manning Publications, 2015.

MICROSOFT CORPORATION. **"Transact-SQL Reference"**. Documentação Oficial da Microsoft.

MITCHELL, T. M. **Machine Learning**. McGraw-Hill, 1997.

MISRA, H.; RAHMAN, H. **Managing Enterprise Information Technology Acquisitions: Assessing Organizational Preparedness**. IGI Global, 2013.

OLIVEIRA, L. B. **Conceitos Gerais de Business Intelligence**. Amazon, 2019.

PONNIAH, P. **Data Warehousing Fundamentals for IT Professionals**. New Jersey. John Wiley & Sons, Inc, 2010.

RIKHARDSSON, PALL; YIGITBASIOGLU, OGAN. **"Business Intelligence & Analytics In Management Accounting Research: Status And Future Focus"**. International Journal of Accounting Information Systems, Elsevier, 2018.

SHARDA, R., DELEN, D., TURBAN, E.; ARONSON, J. E. **Business Intelligence and Analytics: Systems for Decision Support**. Pearson, 2014.

SILBERSCHATZ, A., KORTH, H. F.; SUDARSHAN, S. **Database System Concepts**. McGraw-Hill, 2010.

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



TURBAN, E., KING, D., MCKAY, J., LEE, J., D., V.; MARSHALL, P. **Eletronic Commerce: A Managerial Perspective**. Pearson Education Inc., Upper Saddle River, 2008.