**Implantação de ELT como alternativa ao ETL para transformação e análise de dados em grandes empresas1**

Adson Barbosa de Souza2

**Resumo**

Com a evolução da computação em nuvem, há um crescente aumento da quantidade de dados gerados em diversas áreas, esse grande volume de dados tem potencial para fornecer insights e melhorar a tomada de decisão. Para o tratamento desses dados, são necessários aplicar métodos de extração e transformação, dois desses métodos são o ETL e o ELT, que tem como função transformar e disponibilizar esses dados em sistemas de armazenamento como o Data Warehouse e o Data Lake. Para que assim esses dados sejam disponibilizados em ferramentas de BI para que seja realizada a tomada de decisão.

**Palavras-chave:** Big Data; ETL; ELT; Data Warehouse; Data Lake.

**Introdução**

Nos últimos anos, temos observado uma explosão na quantidade de dados gerados em diversas áreas e setores. Esse grande volume de informações, conhecido como *Big Data*, tem potencial para fornecer insights valiosos e melhorar a tomada de decisão em organizações. No entanto, para que os dados sejam úteis, é preciso que eles sejam coletados, armazenados e processados de maneira adequada (SILVA,2022).

Segundo (Rautenberg e Carmo, 2019):

“Big Data se trata de um conjunto de dados impeditivo de captura, armazenamento, gerenciamento e análise por parte de ferramentas computacionais tradicionais, requer formas inovadoras de processamento de grandes volumes de dados heterogêneos[...]”.

Para lidar com esses desafios, surgiram tecnologias como *Data Warehouses, Data Lakes* e processos de extração como, *Extract, Transform, Load* (ETL) e *Extract, Load, Transform* (ELT), que permitem capturar, armazenar e processar grandes volumes de dados para análises (MICROSOFT, 2023).

(Bansal e Kagemann, 2015), eles decrevem:

“[...] A transformação inclui limpeza, racionalização e complementação dos registros, o processo de limpeza removerá erros e padronizará as informações e a complementação implicará no acréscimo de dados[...]”.

A escolha entre a ETL e ELT para a transformação dos dados, pode ter um impacto significativo no sucesso de um projeto de integração de dados. Compreender os fundamentos teóricos e as aplicações práticas dessas duas abordagens é essencial para tomar uma decisão informada (FÁTIMA, 2020).

**Objetivos**

Comparar os métodos de extração de dados ETL e ELT. A fim de definir qual o melhor método a ser utilizado, levando em consideração suas vantagens e desvantagens para cada cenário específico encontrado. Identificar os fatores que influenciam a escolha entre ETL e o ELT.

**Problema**

As empresas estão enfrentando desafios cada vez maiores na gestão e análise de grandes volumes de dados devido ao aumento exponencial da coleta de dados. As soluções tradicionais de ETL e *Data Warehouse* foram amplamente utilizadas, mas estão começando a enfrentar limitações em termos de escalabilidade, complexidade, custos e flexibilidade. Como alternativa, a implantação de uma solução ELT e o uso de um *Data Lake* podem oferecer vantagens significativas para as empresas que precisam gerenciar e analisar grandes quantidades de dados não estruturados. No entanto, ainda há poucos estudos que exploram os benefícios e desafios dessas abordagens alternativas (GARCIA, 2022).

**Metodologia**

Esta pesquisa foi realizada por meio de uma revisão bibliográfica da literatura com o objetivo de comparar os métodos de extração de dados ETL e ELT.

A análise dos dados incluiu uma revisão crítica e síntese dos resultados encontrados em diferentes fontes, a comparação dos resultados de diferentes autores.

Os resultados e conclusões foram apresentados por meio de uma análise crítica dos materiais selecionados. Foram identificadas as principais características dos métodos ETL e ELT, bem como suas diferenças e semelhanças. Também foi apresentada uma análise crítica das vantagens e desvantagens de cada método, além de uma reflexão sobre a relevância desses resultados e suas implicações para a prática profissional.

**Resultados**

O conceito de *Big Data* pode ser condensado com sendo um grande volume dados, sendo eles estruturados ou não estruturados, provenientes de diversas fontes, que devem ser gerenciados e analisados, o maior desafio é transformar todo o volume de dados em informação, gerando assim conhecimento e valor para as organizações (JUNIOR, 2016).

O conceito de *Big Data*, já foi dividido em 3 Vs (volume, velocidade e variedade), com o passar do tempo os conceitos foram atualizados e hoje já pode-se encontrar referências que abordam o conceito com 5 Vs (volume, velocidade, variedade, valor e veracidade) (ARORA, 2021).

*Data Warehouse* é um repositório que armazena dados estruturados e semiestruturados para relatórios e análises. Assim como o *Data Lake* o *Data Warehouse* pode armazenar grandes quantidades de informações e ajuda as empresas na tomada de decisão (FÁTIMA, 2019).

*Data Lake* é um repositório que armazena todos os dados estruturados, não estruturados e semiestruturados. Ajuda muitas empresas na tomada de decisão, as empresas que implementaram os *Data Lakes* superaram 9% a performance de empresas semelhantes no crescimento orgânico da receita (AWS, 2023).

O ETL é baseado na teoria de que a integração dos dados deve ser executada antes que os dados sejam carregados em um sistema de destino para garantir a consistência e a precisão dos dados. Em contraste, o ELT, é baseado na teoria de que os dados devem ser carregados em um sistema destino primeiro e depois transformados conforme necessário. É importante destacar que os processos de ETL e ELT, apesar de diferentes, solucionam o mesmo problema (BLASI, 2020).

O ETL é um pipeline de dados usados para coletar dados de várias fontes, após a coleta vem a transformação dos dados de acordo com as regras de negócios e o carregamento dos dados em um armazenamento de dados de destino (MICROSOFT, 2023).

O ELT é um pipeline que a transformação ocorre no armazenamento de dados de destino, sem ter a necessidade de usar um mecanismo de transformação separado (MICROSOFT, 2023).

O processo de ELT se tornou popular com a adoção da infraestrutura em nuvem, que oferece aos bancos de dados de destino o poder de processamento necessário para as transformações (AWS, 2023).

**Conclusão**

O ETL se tornou uma parte essencial nos processos de inteligência de negócios, possibilitando assim que dados brutos de diferentes fontes sejam integrados em um único local, para que sejam extraídas as informações necessárias para o negócio. Ao contrário do ETL o ELT, os dados serão integrados primeiramente para que depois sejam convertidos, reduzindo consideravelmente o tempo de carregamento e sendo assim um método mais eficiente em termos de recursos (FÁTIMA, 2020).

O ETL se tornou uma parte essencial nos processos de inteligência de negócios, possibilitando assim que dados brutos de diferentes fontes sejam integrados em um único local, para que sejam extraídas as informações necessárias para o negócio. Ao contrário do ETL o ELT, os dados serão integrados primeiramente para que depois sejam convertidos, reduzindo consideravelmente o tempo de carregamento e sendo assim um método mais eficiente em termos de recursos (FÁTIMA, 2020).

**Referências**

ARORA, Vishesh. **O que é Big Data?**: introdução, usos e aplicações.. Introdução, Usos e Aplicações.. 2021. Disponível em: https://www.analyticsvidhya.com/blog/2021/05/what-is-big-data-introduction-uses-and-applications/. Acesso em: 24 mar. 2022.

AWS. **O que é ETL?** Disponível em: https://aws.amazon.com/pt/what-is/etl/. Acesso em: 11 abr. 2023.

AWS. **O que é um data lake?**: armazene todos os seus dados em um repositório centralizado em qualquer escala. Armazene todos os seus dados em um repositório centralizado em qualquer escala. Disponível em: https://aws.amazon.com/pt/big-data/datalakes-and-analytics/what-is-a-data-lake/. Acesso em: 04 abr. 2023.

BANSAL, Srividya K. Towards a Semantic Extract-Transform-Load (ETL) Framework for Big Data Integration. 2014 IEEE International Congress on Big Data, [s. l.], 2014.

BLASI, Isabela. **ETL X ELT: qual a diferença?** 2020. Disponível em: https://blog.indicium.tech/etl-vs-elt-diferencas/?utm\_source=Google&utm\_medium=cpc&utm\_term=&utm\_campaign=19229929630&utm\_content=&gclid=Cj0KCQiAxbefBhDfARIsAL4XLRpHfjmOB5-JAR-YWRjnhhC1TvnUQNWYbNgtbLE8Rzm5dxDS9LglxFwaAtQCEALw\_wcB. Acesso em: 14 mar. 2023.

FÁTIMA, Nida. **ETL vs. ELT: Qual é a diferença?** 2020. Disponível em: https://www.astera.com/pt/type/blog/etl-vs-elt-whats-the-difference/. Acesso em: 14 mar. 2023.

FÁTIMA, Nida. **O que é um data warehouse**: definição, exemplo e benefícios. definição, exemplo e benefícios. 2019. Disponível em: https://www.astera.com/pt/tipo/blog/defini%C3%A7%C3%A3o-de-data-warehouse/. Acesso em: 04 abr. 2023.

GARCIA, Marco. **Big Data**: o que é, conceito e definição. O que é, conceito e definição. 2022. Disponível em: https://cetax.com.br/big-data/. Acesso em: 22 mar. 2023.

MICROSOFT. **Extract, transform, and load (ETL)**. Disponível em: https://learn.microsoft.com/en-us/azure/architecture/data-guide/relational-data/etl. Acesso em: 11 abr. 2023.

MICROSOFT. **O que é Data Lake**: veja como os data lakes diferem de data warehouses e data lakehouses. descubra como criar uma base escalonável para todas as suas análises com o azure. Disponível em: https://azure.microsoft.com/pt-br/resources/cloud-computing-dictionary/what-is-a-data-lake/#what-is-a-data-lake. Acesso em: 03 abr. 2023

MICROSOFT. **O que é um data warehouse?**: saiba o que é data warehouse, os benefícios de usar um, as práticas recomendadas a serem consideradas durante a fase de design e quais ferramentas incorporar quando finalmente for a hora de criar. Disponível em: https://azure.microsoft.com/pt-br/resources/cloud-computing-dictionary/what-is-a-data-warehouse/#get-started. Acesso em: 04 abr. 2023.

RAUTENBERG, Sandro; CARMO, Paulo Ricardo Viviurka do. Big data e ciência de dados. **Brazilian Journal Of Information Science**: research trends, [S.L.], v. 13, n. 1, p. 56-67, 29 mar. 2019. Faculdade de Filosofia e Ciências. <http://dx.doi.org/10.36311/1981-1640.2019.v13n1.06.p56>.

SILVA, Leidiane Angelica Nunes da. **CIÊNCIA DE DADOS COMO MÉTODO DE TRANSFORMAÇÃO DE DADOS EM INFORMAÇÃO**. 2022. 24 f. TCC (Graduação) - Curso de Sistemas Para Internet, Instituto Federal de Educação Ciência e Tecnologia, Salgueiro, 2022.