

**UNIAESO - Centro Universitário AESO-Barros Melo**  
**Curso de Bacharelado em Sistemas de Informação**

**ADSON BARBOSA DE SOUZA**

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

**Olinda**  
**2023**

**ADSON BARBOSA DE SOUZA**

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

Trabalho de Conclusão de Curso  
apresentado ao Curso Bacharelado em  
Sistemas de Informação, UNIAESO -  
Centro Universitário AESO-Barros Melo.  
Orientador(a): Msc Ameliara Freire  
Santos de Miranda.

**Olinda  
2023**

## SUMÁRIO

1 INTRODUÇÃO.....	4
1.1 OBJETIVOS.....	5
1.1.1 Objetivo Geral.....	5
1.1.2 Objetivos Específicos.....	5
1.2 PROBLEMA.....	5
1.3 METODOLOGIA.....	6
2 FUNDAMENTAÇÃO TEÓRICA.....	6
2.1 BIG DATA.....	6
2.1.1 Os 5 Vs.....	7
2.2 DATA WAREHOUSE.....	9
2.2.1 BENEFÍCIOS DO DATA WAREHOUSE.....	9
2.2.2 ARQUITETURA DO DATA WAREHOUSE.....	10
2.2.3 PRINCIPAIS CARACTERÍSTICAS DE UM DATA WAREHOUSE.....	10
2.2.4 BENEFÍCIOS DO DATA WAREHOUSE.....	10
2.3 DATA LAKE.....	11
2.3.1 ELEMENTOS ESSENCIAIS DE UMA SOUÇÃO DE DATA LAKE.....	11
2.3.1.1 MOVIMENTAÇÃO DE DADOS.....	11
2.3.1.2 ARMAZENE E CATALOGUE DADOS COM SEGURANÇA.....	11
2.3.1.3 ANÁLISE DE DADOS.....	12
2.3.1.4 MACHINE LEARNING.....	12
2.3.2 EXEMPLOS DE ONDE OS DATA LAKES AGREGAM VALORES.....	12

2.3.2.1 MELHORES INTERAÇÕES COM O CLIENTE.....	12
2.3.2.2 AUMENTO DA EFICIÊNCIA OPERACIONAL.....	12
2.3.3 A ARQUITETURA DO DATA LAKE.....	13
2.4 ETL.....	14
2.4.1 Processo de ETL.....	14
2.4.2 COMO FUNCIONA A EXTRAÇÃO DE DADOS.....	14
2.4.3 COMO FUNCIONA A TRANSFORMAÇÃO DE DADOS.....	15
2.4.4 COMO FUNCIONA O CARREGAMENTO DE DADOS.....	15
2.5 ELT.....	16
2.5.1 PROCESSO DE ELT.....	16
4 ETL X ELT.....	17
4.1.1 Vantagens ETL.....	18
4.1.2 Desvantagens ETL.....	18
4.1.3 Vantagens ELT.....	19
4.1.4 Desvantagens ELT.....	19
5 CONCLUSÃO.....	20
6 REFERÊNCIAS.....	22

## 1 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).

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

Os *Data Warehouses* são sistemas de armazenamento projetados para suportar a análise de grandes quantidades de dados estruturados. Já os *Data Lakes* são sistemas mais flexíveis e escaláveis, que permitem armazenar dados de diferentes fontes e formatos, sem a necessidade de transformá-los em um formato estruturado antes de serem armazenados. Os processos de ETL e ELT, por sua vez, são fundamentais para carregar e transformar dados nos *Data Warehouses* e *Data Lakes* (AWS, 2023).

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[...]”.

O ETL é um processo tradicional de transformação de dados formado por três etapas: a extração, a transformação e o carregamento de dados. A principal vantagem do ETL, é que ele permite a criação de uma *Data Warehouse* unificado, que pode ser usado para relatórios e análises (BLASI, 2020).

O ELT pode ser considerado a modernização do processo de ETL. Ao contrário do ETL, o ELT é um processo mais ágil para o carregamento e o processamento de dados. A sua principal vantagem é que ele permite um processamento de dados mais rápido, pois esses dados são carregados no sistema de destino antes de serem transformados (BLASI, 2020).

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).

## **1.1 OBJETIVOS**

### **1.1.1 OBJETIVO GERAL**

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

### **1.1.2 OBJETIVOS ESPECÍFICOS**

- Examinar as principais diferenças entre o processo de ETL em relação ao processo de ELT e como elas afetam o desempenho, a escalabilidade e a flexibilidade da integração de dados.
- Identificar os fatores que influenciam a escolha entre ETL em relação ao ELT, como volume, estrutura e complexidade dos dados.

## **1.2 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).

### 1.3 METODOLOGIA

Esta pesquisa foi realizada por meio de uma revisão sistemática 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.

## 2 FUNDAMENTAÇÃO TEÓRICA

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).

### 2.1 *BIG DATA*

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).

Arrais (2022) descreve:

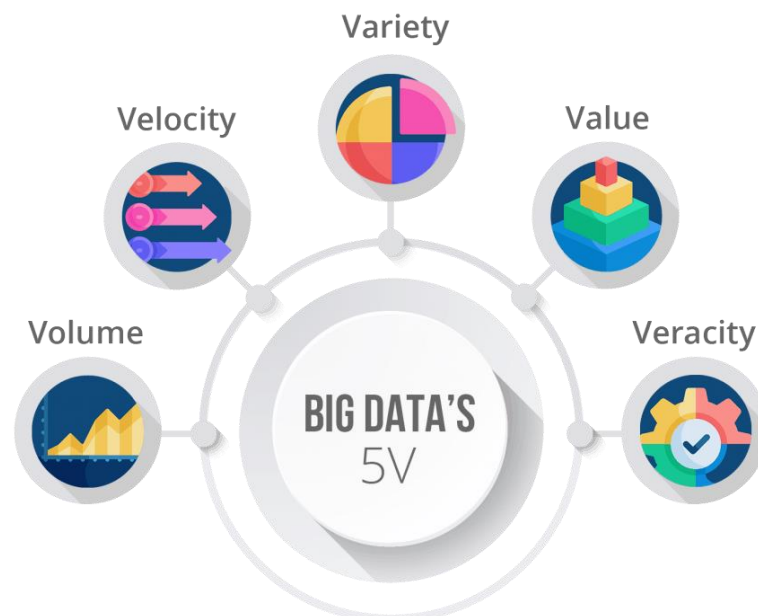
“[...] Temos sensores em todos os lugares, crescimento da utilização de dispositivos conectados à rede, aumento do poder de armazenamento com um arquivamento quase infinito, nuvens de processadores com evoluções em todos os recursos computacionais. Todas essas habilidades em constante evolução resultam em uma geração massiva de dados, que vem

mudando a ciência, a medicina, os negócios e a tecnologia. Todas essas informações, com grande volume e variedade, necessitam de formas inovadoras para ingestão, transformação, armazenamento e análise, pensando na riqueza que esses dados podem conter para a tomada de decisões.

Big Data é um conjunto de dados que devido a quantidade e a variedade não podem ser manipulados com as ferramentas computacionais tradicionais. Temos, portanto, um crescimento exponencial de dados, heterogêneos, oriundos de diferentes fontes, de forma distribuída e descentralizada. [...]”.

### 2.1.1 Os 5 Vs

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 (ARORA, 2021).



Fonte: Arora (2021). Acesso em: 24 mar. 23.

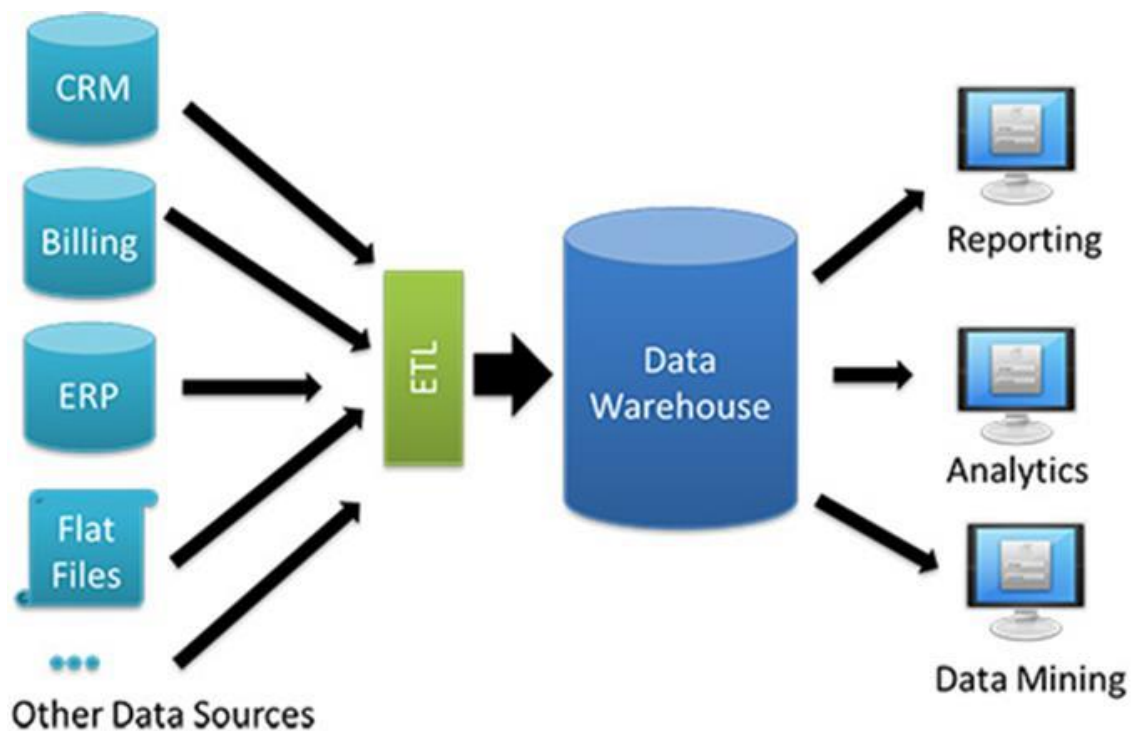


Neste trabalho são utilizados os conceitos descritos por Arrais (2022):

- Volume: *Big Data* é formado por um conjunto de dados tão grande que é necessário ferramentas e estruturas especializadas para seu armazenamento, processamento e análise. Esse grande volume é gerado pelas indústrias, saúde, Internet das Coisas e demais sistemas de forma exponencial. Não existe um limite fixo para que uma quantidade de dados seja considerada como big data, mas normalmente são dados em grande escala que oferecem desafios para armazenamento, gerenciamento e processamento, necessitando de ferramentas não tradicionais.
- Velocidade: Este conceito está relacionado com a rapidez em que os dados são gerados, o que influencia diretamente no crescimento exponencial dos dados e no volume alto para armazenamento. Algumas aplicações precisam analisar esses dados em tempo real, e, portanto, é necessário ter ferramentas especializadas para essa ingestão e análise com alta velocidade para decisões mais assertivas.
- Variedade: Refere-se à variedade de dados que são gerados (dados estruturados, não estruturados e semiestruturados). Temos dados em formato de textos, imagens, áudios, vídeos, oriundos de sensores, de operações; e os sistemas precisam ser flexíveis para atender todos esses formatos e suas especificidades para que possa ser possível o tratamento e aquisição de informações a partir dos mesmos.
- Veracidade: Este conceito se refere ao aspecto de confiabilidade dos dados. Para extrair valor, gerando informações de qualidade, é necessário que os dados sejam limpos e que ruídos (dados incorretos e/ou faltantes sejam identificados e/ou eliminados).
- Valor: Refere-se à utilidade que os dados possuem para a finalidade pretendida, ou seja, para atender alguma necessidade ou resolver um problema. Esse valor está associado diretamente à veracidade e precisão dessas informações, podendo também em algumas aplicações depender da velocidade de processamento para tomada de decisões.

## 2.2 DATA WAREHOUSE

*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).



Fonte: Fátima (2019). Acesso em: 04 abr. 23.

### 2.2.1 Benefícios do *Data Warehouse*

De acordo com Microsoft (2023):

- Consolidação de dados de várias fontes em uma única fonte de verdade.
- Armazenamento e análise de dados históricos de longo prazo abrangendo meses e anos.
- Limpando e transformando dados para que eles sejam precisos, consistentes e padronizados em estrutura e forma.
- Reduzindo os tempos de consulta ao coletar dados e processar análises, o que melhora o desempenho geral em todos os sistemas.
- Carregando dados com eficiência sem precisar lidar com os custos de implantação ou infraestrutura.
- Protegendo dados para que eles sejam privados, protegidos e seguros.

- Preparação de dados para análise por meio de mineração de dados, ferramentas de visualização e outras formas de análise avançada.

### 2.2.2 Arquitetura do *Data Warehouse*

Um *Data Warehouse*, pode ser considerado mais do que apenas um repositório de dados. Ele é um sistema altamente estruturado e cuidadosamente arquitetado composto por várias camadas que interage com seus dados e entre si de maneira diferente, essas camadas incluem Microsoft (2023):

- Camada inferior: Com o processo de extração, transformação e carregamento chamado ELT, responsável também pelos dados armazenados e otimizados onde o tempo de consulta é mais rápido tem melhor desempenho.
- Camada intermediária: Com o processamento analítico online chamado como servidor OLAP, que acessam grandes volumes de dados do *Data Warehouse* onde os resultados são extremamente rápidos.
- Camada superior: É onde apresenta visualmente os dados processados e é a camada mais utilizada pelos analistas que têm acesso e pode usar para a necessidade de relatórios e *Business Intelligence* (BI) de autoatendimento.

### 2.2.3 Principais características de um *Data Warehouse*

De acordo com (ROCK CONTENT, 2021):

- Gestão dos dados
- Integração dos dados
- Não volatilidade dos dados

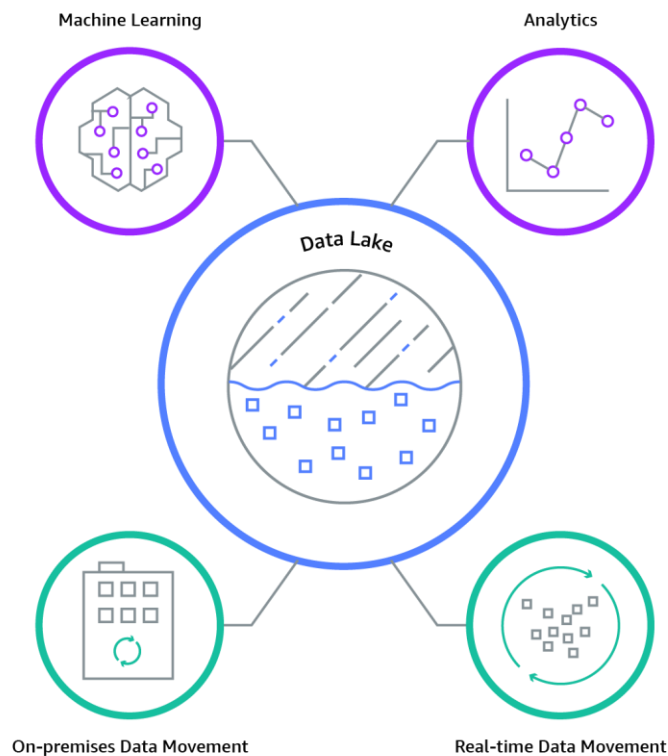
### 2.2.4 Benefícios do *Data Warehouse*

De acordo com (ROCK CONTENT, 2021):

- Agilidade nas consultas
- Dados de qualidade
- Segurança
- Visão histórica
- Escalabilidade
- Autonomia para os funcionários

## 2.3 DATA LAKE

*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).



Fonte: AWS (2023). Acesso em 03 abr. 23.

### 2.3.1 Elementos essenciais de uma solução de *Data Lake*

As empresas que estão criando *Data Lakes* e uma plataforma de análise precisam de vários recursos importante como:

#### 2.3.1.1 Movimentação de dados

Os *Data Lakes* permitem importar qualquer quantidade de dados em tempo real economizando tempo na definição de estrutura de dados, esquemas e transformação (AWS, 2023).

#### 2.3.1.2 Armazene e catalogue dados com segurança

Os *Datas Lakes* permitem armazenar dados relacionais, como banco de dados operacionais e dados de aplicações de linha de negócios, dados não relacionais também como aplicativos móveis, dispositivos *Internet of Things* (IoT) e mídias

sociais. Tem capacidade de atender dados por meio de *crawling*, catalogação e indexação de dados. Os dados têm que ser protegidos para garantir que os ativos dados sejam protegidos (AWS, 2023).

#### 2.3.1.3 Análise de Dados

Os *Data Lakes* permitem executar análises sem a necessidade de mover os dados para um sistema de análise separado. Permitindo ter várias funções dentro de uma empresa, como cientistas de dados, desenvolvedores de dados e analistas de negócios, acessando dados de ferramentas e *frameworks* analíticos. Incluindo *frameworks* de código aberto, como Apache Hadoop, Presto, Apache Spark e ofertas comerciais de fornecedores de *Data Warehouse* e inteligência empresarial (AWS, 2023).

#### 2.3.1.4 Machine Learning

O *Data Lake* permite que a empresa gere diferentes tipos de situações, incluindo relatórios de dados históricos e *Machine Learning*, criando modelos para prever resultados sugerindo uma série de ações prescritas para chegar no resultado ideal (AWS, 2023).

### 2.3.2 Exemplos de onde os *Data Lakes* agregam valores

#### 2.3.2.1 Melhores interações com o cliente

O *Data Lake* ajuda na combinação de dados do cliente em uma plataforma *Customer Relationship Management* (CRM) com análise de mídia social, uma plataforma de marketing para capacitar a empresa a entender o grupo de clientes mais lucrativos e a causa da perda de cliente, as promoções e recompensas que aumentam a fidelidade (AWS, 2023).

#### 2.3.2.2 Aumento da eficiência operacional

O *Data Lake* facilita o armazenamento e execução de análise em dados de IoT gerados por máquina para descobrir soluções para reduzir custos operacionais e aumentar a qualidade. A Internet das Coisas (IoT) apresenta várias maneiras de coletar dados sobre o processo, como fabricação de dados em tempo real com dispositivos conectados à internet (AWS, 2023).

Os desafios dos *Data Lakes* são os dados brutos que são armazenados sem supervisão do conteúdo. O *Data Lake* torna os dados utilizáveis, precisa de um

mecanismo definido para catálogo e proteger os dados, sem isso os dados não podem ser encontrados ou serem confiáveis. Para atender as necessidades do público mais amplo o *Data Lake* tem que ter governança, consistência semântica e controle de acesso (AWS, 2023).

O *Data Lake* é ideal para ser implantado na nuvem, pois oferece uma performance, escalabilidade, confiabilidade, disponibilidade, mecanismo analítico e enormes economias. Os principais motivos que os clientes perceberam a nuvem como vantagem para *Data Lakes* são: segurança, implantação mais rápida, disponibilidade, atualização de recursos, funcionalidades, elasticidade, cobertura geográfica e custo vinculados à utilização real (AWS, 2023).

### 2.3.3 A arquitetura do *Data Lake*

O armazenamento e o processamento de dados trabalham juntos para criar uma arquitetura em camadas coesivamente, que é informada por *Big Data* e executada sobre o *Data Lake*. A arquitetura pode formar uma estrutura operacional de um *Data Lakehouse*. Cada empresa tem uma configuração própria e exclusiva, porém a maioria das arquiteturas de *Data Lakehouse* tem o seguinte (AWS, 2023):

- Gerenciamento e orquestração de recursos
- Conectores para fácil acesso
- Análise confiável
- Classificação de dados
- Extrair, carregar, transformar (ELT) processos
- Segurança e suporte
- Governança e administração

## 2.4 ETL

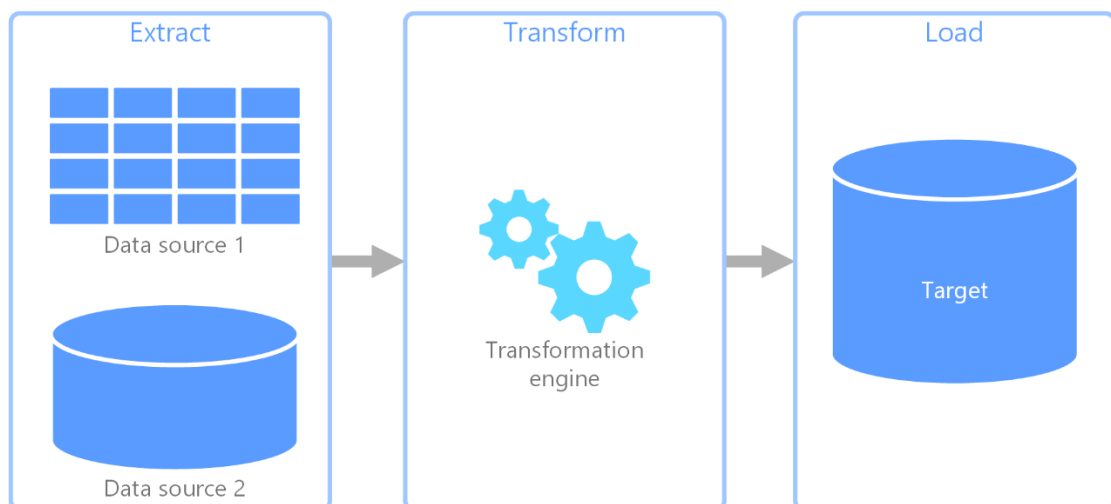
É um pipeline de dados usados para coletar dados de várias fontes, após a coleta vem a transformação dos dados de acordo com as regras de negócios e o carregamento dos dados em um armazenamento de dados de destino. A transformação de dados que ocorre geralmente envolve várias operações, como filtragem, classificação, agregação, junção de dados, limpeza de dados, deduplicação e validação de dados. As três fases do ETL são executadas em paralelo para economizar tempo (MICROSOFT, 2023).

### 2.4.1 Processo de ETL

Processo tradicional de transformação de dados formado por três etapas: a extração, a transformação e o carregamento de dados (BLASI, 2020).

De acordo com (AWS, 2023):

- Extração de dados relevante do banco de dados de origem
- Transformação dos dados para que sejam mais adequados a análises.
- Carregamento dos dados no banco de dados de destino.



Fonte: Microsoft (2023), Acesso em 11 abr. 23.

### 2.4.2 Como funciona a extração de dados

A extração de dados, copiam ou extraem dados brutos de várias fontes e armazenam na área de preparação onde os dados extraídos são armazenados temporariamente, isso significa que o conteúdo armazenado é excluído após a

conclusão da extração dos dados. A área de preparação pode reter um arquivo de dados para fins de solução de problemas (AWS, 2023).

A extração de dados acontece de uma das três maneiras:

- Notificação de atualização
- Extração gradual
- Extração completa

#### 2.4.3 Como funciona a transformação de dados

A transformação de dados consolida e transforma os dados brutos na área de preparação onde vão ser preparados para o *Data Warehouse* de destino. A transformação de dados pode envolver os seguintes tipos de alteração de dados (AWS, 2023):

- Transformação de dados básica (AWS, 2023)
  - Limpeza de dados
  - Eliminação de duplicação de dados
  - Revisão de formato de dados
- Transformação de dados avançada (AWS, 2023)
  - Derivação
  - Junção
  - Separação
  - Resumo
  - Encriptação

#### 2.4.4 Como funciona o carregamento de dados

O carregamento de dados, movem os dados transformados da área de preparação para o data warehouse de destino (AWS, 2023).

O carregamento de dados tem dois métodos (AWS, 2023):

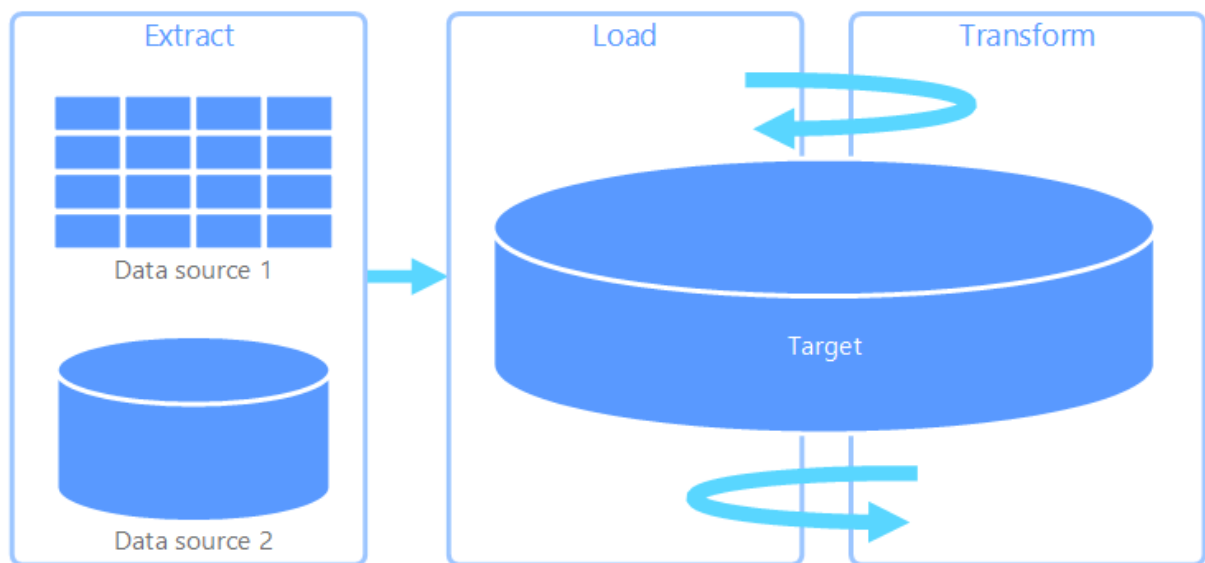
- Carregamento completo
- Carregamento Incremental
  - Carregamento incremental por transmissão
  - Carregamento incremental em lotes



## 2.5 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. Os recursos de processamento do armazenamento de dados de destino são usados para transformar os dados (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).



Fonte: Microsoft (2023). Acesso em 11 abr. 23.

### 2.5.1 Processo de ELT

Processo mais ágil para o carregamento e o processamento de dados. A sua principal vantagem é que ele permite um processamento de dados mais rápido, pois esses dados são carregados no sistema de destino antes de serem transformados (BLASI, 2020).

De acordo com Blasi (2020):

- **Extração:** Faz extração de dados brutos de diversas fontes para a integração em um repositório de dados único (BLASI, 2020).
- **Carregamento:** O carregamento dos dados coletados em um *Data Warehouse* ou repositório de dados (BLASI, 2020).

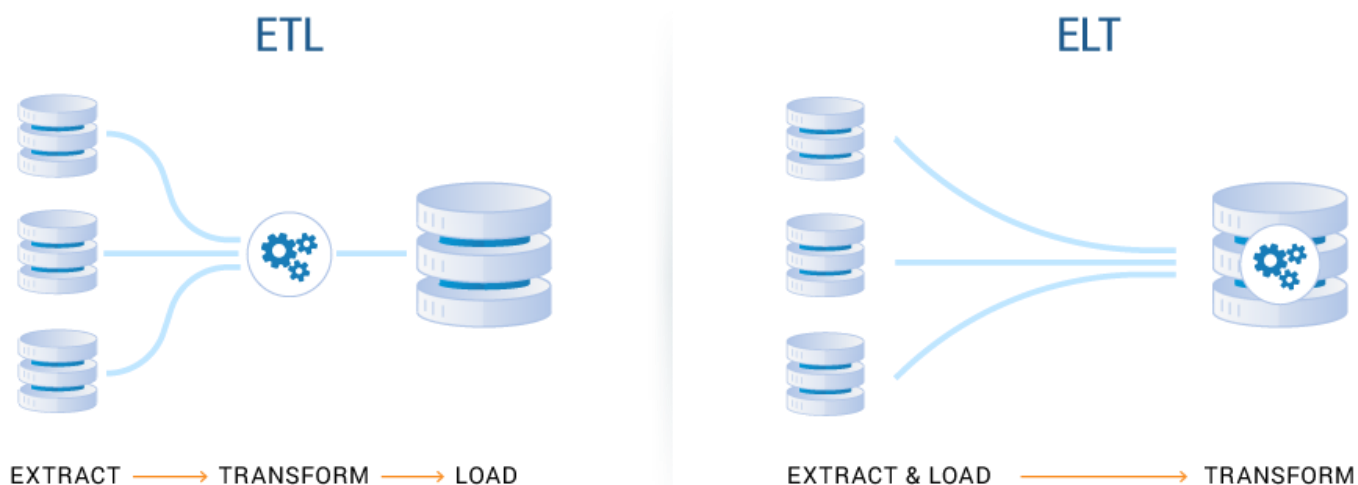
- Transformação: Na transformação dos dados brutos em dados modelados dentro de um *Data Warehouse* para a aplicação de *Business Intelligence*, análise de dados e *advanced analytics*.

#### 4 ETL x ELT (Comparativo)

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 ELT possui muitos benefícios em relação a escalabilidade e em relação ao processamento e armazenamento em nuvem, cada abordagem oferece vantagens e desvantagens a serem consideradas. A velocidade do ELT, que vai oferecer a possibilidade de carregamento e transformação simultâneos, é um fator predominante em seu favor, se você precisar carregar e analisar grandes quantidades de dados mantendo os dados brutos para futuras análises. Se esse os dados coletados e armazenados, são de um modelo específico que suas fontes raramente possam variar, o ETL será uma melhor opção (SEGNER, 2023).

Cada abordagem oferece vantagens e desvantagens a serem consideradas.



Fonte Pauncz (2021). Acesso em 24/04/2023

#### 4.1 Vantagens ETL

- O ETL equilibra a capacidade e compartilhar o trabalho com o sistema de gerenciamento de banco de dados relacional (FÁTIMA, 2020).
- Ao utilizar mapas de dados, o ETL executa operações mais complexas em um único diagrama de fluxo de dados (FÁTIMA, 2020).
- Pode lidar com segregação e paralelismo (FÁTIMA, 2020).
- Processa os dados transmitidos da origem e carrega no destino em lotes (FÁTIMA, 2020).
- Preserva as plataformas de fonte de dados atuais sem se preocupar com a sincronização de dados (FÁTIMA, 2020).
- No processo de ETL de *Business Intelligence* (BI), o processamento das informações é realizado linha a linha, devido a isso ele possui uma boa integração com sistemas de terceiros (FÁTIMA, 2020).
- O ETL transfere apenas dados que já foram transformados, com isso pode haver uma economia nos custos de armazenamento armazenando apenas os dados necessários (PAUNCZ, 2021).
- Suporta gerenciamento de dados complexos, ajudando com cálculos, integrações de dados e manipulação de *strings* (IBM, 2021).
- Segurança e conformidade na criptografia de dados confidenciais (IBM, 2021).

#### 4.2 Desvantagens ETL

- O ETL necessitará de despesas adicionais de *hardware* (FÁTIMA, 2020).
- O processamento das informações linha a linha poderá diminuir o desempenho do processo de ETL (FÁTIMA, 2020).
- Flexibilidade reduzida devido à dependência de fornecedores de ferramentas (FÁTIMA, 2020).
- Os dados necessitam ser transferidos para uma camada adicional antes de chegarem ao destino (FÁTIMA, 2020).
- Não existem controle programado de erros ou mecanismo de recuperação (FÁTIMA, 2020).
- O processo necessita de atualizações periódicas ao invés de atualizações em tempo real (IBM, 2021).

- Tempo mais longo de carregamento devido ao grande número de etapas no estágio de transformação (IBM, 2021).
- Maior tempo de desenvolvimento (IBM, 2021).

#### 4.3 Vantagens ELT

- Melhor desempenho e segurança de dados (FÁTIMA, 2020).
- Necessita de menos tempo e recursos, pois os dados serão transformados e carregados em paralelo (FÁTIMA, 2020).
- Permite dados de tamanho maiores (FÁTIMA, 2020).
- Não necessita de um bloco de transformação, pois o sistema de destino que realiza esse trabalho (FÁTIMA, 2020).
- Maior flexibilidade, não exigindo o desenvolvimento de pipelines complexos (PAUNCZ, 2021).
- A transformação de dados é realizada apenas para os dados necessários para que seja realizada uma análise específica (PAUNCZ, 2021).
- Implementação mais rápida (IBM, 2021).
- Utilização de sistemas de armazenamento em nuvem (IBM, 2021).
- Suporte a tipos de dados estruturados, não estruturados, semiestruturados e brutos (IBM, 2021).

#### 4.4 Desvantagens ELT

- Limitação nas ferramentas que oferecem suporte para o ELT (FÁTIMA, 2020).
- Falta de modularidade devido ao design baseado em conjunto (FÁTIMA, 2020).
- O carregamento de dados confidenciais antes de transformá-los, expõe os dados privados em logs acessíveis aos administradores do sistema (PAUNCZ, 2021).
- A evolução recente causa uma falta de confiança nas ferramentas de ELT em relação as de ETL (PAUNCZ, 2021).
- Demanda uma maior quantidade de armazenamento (BLASI, 2020).

## 5 Conclusão

Quando a abordagem de ETL foi criada, o custo do armazenamento de dados em repositórios era muito elevado, além desse custo o seu processamento era extremamente lento e devido a isso não havia tanto investimento nessas tecnologias. Com a inclusão da computação na nuvem, os processos ganharam velocidade e tiveram os seus custos reduzidos. Considerando o grande volume de dados para a serem processados, muitas empresas têm optado pelo ELT em relação ao ETL, buscando uma maior flexibilidade e agilidade (BLASI, 2020).

O processo de ELT é mais acessível, requer menos tempo e menos recursos, entretanto, se o sistema de armazenamento de destino não for robusto o bastante para o ELT, a escolha pelo ETL será mais apropriada (PAUNCZ, 2021).

O ELT será mais bem aproveitado em conjuntos de dados de alto volume ou em ambientes de uso de dados em tempo real. Enquanto o ETL é mais bem utilizado na sincronização de vários ambientes de dados e na migração de sistemas legados (IBM, 2021).

Para transformações que sejam intensivas e que não há um alto grau de variabilidade, dados que contenham informações pessoais ou sistemas legados, o ETL continua sendo a principal escolha, diante da recente evolução do ELT (SEGNER, 2023).

Principais diferenças:

	ETL	ELT
<b>Extração</b>	Dados brutos extraídos de fontes diferentes	Dados brutos extraídos de fontes diferentes
<b>Transformação</b>	Dados brutos são transformados em servidor secundário ou área de preparação	Os dados são transformados dentro do data warehouse ou sistema
<b>Carregando</b>	Dados carregados no data warehouse/sistema após a transformação	Os dados brutos são carregados diretamente no data warehouse ou sistema

<b>Tipos de dados</b>	Estruturada	Estruturado, semiestruturado, não estruturado
<b>Volume</b>	Mais adequado para conjuntos de dados menores que requerem transformação complexa	Útil para grandes conjuntos de dados que precisam ser carregados rapidamente
<b>Privacidade</b>	a transformação de pré-carregamento pode abordar questões de privacidade	Requer proteção e segurança adicionais
<b>Compatibilidade do Data Lake</b>	Nega muitas vantagens de flexibilidade	sim

Fonte: Segner (2023), acesso em 24/03/2023.

### 3 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.

ARRAIS, Karolayne Fernandes. **CONSTRUÇÃO DE UM PIPELINE DE DADOS UTILIZANDO SERVIÇOS DA NUVEM**. 2022. 91 f. TCC (Graduação) - Curso de Bacharelado em Engenharia de Computação, Centro de Ciências Exatas e de Tecnologia, Universidade Federal de São Carlos, São Carlos - Sp, 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.

BAHGA, A.; MADISETTI, V. **Big Data Science & Analytics: A Hands-On Approach**. Published by Arshdeep Bahga & Vijay Madisetti. 2019.

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](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.

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.

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.

IBM. **ELT vs. ETL: What's the Difference?** 2021. Disponível em: <https://www.ibm.com/cloud/blog/elt-vs-etl-whats-the-difference>. Acesso em: 14 dez. 2021.

JUNIOR, Jose Carlos Da Silva Freitas; MAÇADA, Antonio Carlos Gastaud; OLIVEIRA, Mirian; BRINKHUES, Rafael Alfonso. BIG DATA E GESTÃO DO CONHECIMENTO: DEFINIÇÕES E DIRECIONAMENTOS DE PESQUISA. **REVISTA ALCANCE**, [S. l.], p. 04-22, 1 nov. 2016.

MAGNUM, Lucas. **Engenharia de Dados — EL, ETL e ELT:** abordagens de extração de dados de forma simplificada. Abordagens de extração de dados de forma simplificada. 2021. Disponível em: <https://lucasmagnum.medium.com/engenharia-de-dados-el-etl-e-elt-b42142058c87>. Acesso em: 14 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.

PAUNCZ, Alex. **ETL vs. ELT: Which is Right for Your Data Warehouse?** 2021. Disponível em: [https://www.cdata.com/blog/20210706-etl-vs-elt?kw=&cpn=2023385644&utm\\_source=google&utm\\_medium=cpc&utm\\_campaign=CData\\_-\\_Search\\_-\\_Branding\\_-\\_DSA&utm\\_content=Branding\\_-\\_DSA&utm\\_term=|&kw=&cpn=2023385644&gclid=Cj0KCQiAxbefBhDfARIsAL4XLRq3JP\\_lo9i\\_9EkvWfp2MfeKGG06tQZZ8Vq7O0HjCcs0lZVzuOso-zlaAirnEALw\\_wcB](https://www.cdata.com/blog/20210706-etl-vs-elt?kw=&cpn=2023385644&utm_source=google&utm_medium=cpc&utm_campaign=CData_-_Search_-_Branding_-_DSA&utm_content=Branding_-_DSA&utm_term=|&kw=&cpn=2023385644&gclid=Cj0KCQiAxbefBhDfARIsAL4XLRq3JP_lo9i_9EkvWfp2MfeKGG06tQZZ8Vq7O0HjCcs0lZVzuOso-zlaAirnEALw_wcB). Acesso em: 14 mar. 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>.

ROCK CONTENT. **O que é um Data Warehouse e quais são as suas principais características?** 2021. Disponível em: <https://rockcontent.com/br/blog/data-warehouse/>. Acesso em: 05 abr. 2023.

SEGNER, Michael. **ETL vs ELT: What's the Difference (and Which is Better)?** 2023. Disponível em: <https://www.montecarlodata.com/blog-etl-vs-elt/>. Acesso em: 01 mar. 2023.

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.