

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

**ADSON BARBOSA DE SOUZA**

Análise comparativa dos métodos de extração de dados ETL e ELT e suas  
vantagens de desvantagens

**Olinda**  
**2023**

**ADSON BARBOSA DE SOUZA**

Análise comparativa dos métodos de extração de dados ETL e ELT e suas vantagens e desvantagens

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**

## RESUMO

Com o crescente aumento da quantidade de dados disponíveis, houve o surgimento do termo *Big Data*, a criação do sistema de armazenamento de dados *Data Warehouse* e os modelos de extração de dados, dentre eles o ETL (*Extract, Transform, Load*), que consiste no processo de extrair os dados de um ponto de origem, realizar os tratamentos necessários e realizar o seu carregamento em um *Data Warehouse* de destino. Com o surgimento do armazenamento em nuvem e a evolução dos dados, foi realizado a criação de um novo sistema de armazenamento o *Data Lake* e a atualização do método de extração de dados, de ETL para ELT (*Extract, Load, Transform*). Essa mudança se deu devido ao fato da necessidade de capturar dados de diversas origens, sejam eles estruturados ou não estruturados. Seguindo esse contexto, o objetivo do presente estudo é construir um entendimento sobre as principais diferenças entre os métodos de extração de dados ETL e ELT, onde serão demonstradas suas principais diferenças, vantagens e desvantagens.

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

## ABSTRACT

With the growing increase in the amount of available data, there was the emergence of Big Data, the creation of the Data Warehouse data storage system and the data extraction models, among them the ETL (Extract, Transform, Load), which consists of the process of extracting data from a point of origin, perform the necessary treatments and perform its loading into a destination Data Warehouse. With the emergence of cloud storage and the evolution of data, a new storage system, the Data Lake, was created, which caused the update of the data extraction method, from ETL to ELT (Extract, Load, Transform). This change was due to the need to capture data from various sources, whether structured or unstructured. Following this context, the objective of this study is to build an understanding about the main differences between ETL and ELT data extraction methods, where their main differences, advantages and disadvantages will be demonstrated.

**Keywords:** Big Data; ETL; ELT; Data Lake; Data Warehouse

## SUMÁRIO

SUMÁRIO.....	4
Índice de Figuras.....	6
Índice de Quadros.....	7
1. Introdução.....	8
1.1. Objetivos.....	10
1.1.1. Objetivo Geral.....	10
1.1.2. Objetivos Específicos.....	10
1.2. Problema.....	10
1.3. Metodologia.....	11
2. Fundamentação Teórica.....	11
2.1. Big Data.....	11
2.1.1. Os 5 Vs.....	13
2.2. Data Warehouse.....	14
2.2.1. Arquitetura do Data Warehouse.....	15
2.2.2. Principais Característica de um Data Warehouse.....	16
2.2.3. Benefícios do Data Warehouse.....	16
2.3. Data Lake.....	17
2.3.1. Elementos Essenciais de uma Solução de Data Lake.....	18
2.3.2. Fases Estratégicas.....	19
2.3.3. A Arquitetura do Data Lake.....	20
2.3.4. Comparação entre Data Warehouse e Data Lake.....	21

2.4. ETL.....	21
2.4.1. Processo de ETL.....	22
2.4.2. Como Funciona a Extração de Dados.....	22
2.4.3. Como Funciona a Transformação de Dados.....	23
2.4.4. Como Funciona o Carregamento de Dados.....	23
2.5. ELT.....	23
2.5.1. Processo de ELT.....	24
2.5.2. Transformação de Dados.....	25
3. Comparativo entre ETL e ELT.....	26
3.1. Vantagens ETL.....	27
3.2. Desvantagens ETL.....	27
3.3. Vantagens ELT.....	27
3.4. Desvantagens ELT.....	28
3.5. Principais Diferenças.....	28
4. Conclusão.....	29
5. Referências.....	31

## Índice de Figuras

Figura 1. Os 5Vs do <i>Big Data</i> .....	12
Figura 2. Funcionamento do <i>Data Warehouse</i> .....	14
Figura 3. Funcionamento do <i>Data Lake</i> .....	17
Figura 4. Processo de ETL.....	20
Figura 5. Processo de ELT.....	22
Figura 6. ETL x ELT.....	23

## Índice de Quadros

Quadro 1. <i>Data Warehouse</i> x <i>Data Lake</i> .....	19
Quadro 2. ETL x ELT.....	25



## 1 INTRODUÇÃO

Nos últimos 20 anos, foi observado uma explosão na quantidade de dados gerados em diversas áreas e setores. De acordo com um relatório da *Internacional Data Corporation* (IDC), em 2011, o termo *Big Data* é usado para descrever grandes conjuntos de dados, geralmente incluindo massa de dados não estruturados que irão precisar de análise, esses dados têm 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; CHEN, MAO, LIU, 2014).

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*. Esses sistemas de armazenamento fazem parte do Big Data, as indústrias recentemente se interessaram pelo seu alto potencial (AWS, 2023; ; CHEN, MAO, LIU, 2014).

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 processamento e a extração desses dados são realizados através dos métodos de extração de dados, como exemplo o ETL e o ELT (RAUTENBERG e CARMO, 2019).

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. Com a crescente evolução da computação em nuvem, foi criado um novo método o ELT que é modernização do ETL, para atender

a processos que necessitam utilizar a nuvem, mudando a ordem dos processos de carregamento e transformação (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).

A transformação dos dados, 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, portanto é necessário se definir qual será o processo de extração de dados utilizado (BANSAL e KAGEMANN, 2015).

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. Os processos de ETL e ELT, tem como a etapa mais difícil a transformação de dados, para realização de operações em lote, os processos em ETL são mais adequados, pois os bancos de dados utilizados por esses processos, não inserem ou atualizam dados rapidamente. Para a utilização do ELT juntamente com banco de dados não relacionais, que por natureza possuem fontes de dados altamente dinâmicas e sem esquema, as decisões de negócios baseadas em dados, devem ser tomadas em um tempo hábil e limitado. Compreender os fundamentos teóricos e as aplicações práticas dessas duas abordagens é essencial para tomar uma decisão informada (FÁTIMA, 2020; SINGHAL, AGGARWAL, 2022).

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.

## 1.1 OBJETIVOS

### 1.1.1 OBJETIVO GERAL

Comparar os métodos de extração de dados ETL e ELT, levando em consideração as suas vantagens e desvantagens para cada cenário encontrado, considerando a arquitetura de dados envolvida no processo. Seja para uma arquitetura de dados que tem como base sistemas legados, fontes de dados monolíticas, cujo suas informações não serão alteradas com frequência ou para a implementação do processo em uma arquitetura voltada para nuvem, utilizando banco de dados não relacionais, como exemplo NoSQL, sistemas de arquivos como Hadoop, Amazon S3 ou Snowflake, que são utilizados para ingestão de fontes de dados dinâmicas e sem um esquema definido. É necessário entender o funcionamento do ETL e ELT para que se possa definir qual o método mais apropriado a ser utilizado.

### 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 bibliográfica com o objetivo de comparar os métodos de extração de dados ETL e ELT.

Foi utilizada como fonte de busca para artigos científicos o portal *IEEE Xplore* (<https://ieeexplore.ieee.org>) e o *Google Acadêmico* (<https://scholar.google.com/>). Para seleção dos artigos foram utilizadas as palavras chaves, ETL, ELT, *Big Data*, *Data Warehouse*, *Data Lake*, Transformação de Dados, Computação na Nuvem. Dentre os artigos encontrados, foram selecionados apenas os que continham como conteúdo, as palavras chaves, seja como tópico principal ou como subtópico da pesquisa.

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

Buscando apresentar as principais características dos métodos de extração de dados e as suas principais vantagens e desvantagens. Além de uma reflexão sobre a aplicação desses métodos baseados nos modelos de armazenamento de dados, *Data Lake* e *Data Warehouse*.

## 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).

Possuímos sensores em todos os lugares, causando, crescimento da utilização de dispositivos conectados à rede, aumento do poder de armazenamento com um arquivamento quase infinito, arquitetura em nuvem, com evolução constante de todos os recursos computacionais. Tudo isso ocasiona uma geração massiva de dados, que tem auxiliado diversas áreas. Todas essas informações precisam ser tratadas, armazenadas e analisadas, com foco na riqueza que esses dados podem conter. O Big Data é um conjunto de dados que devido a quantidade e variedade, não podem ser manipulados através de ferramentas computacionais tradicionais (ARRAIS, 2022).

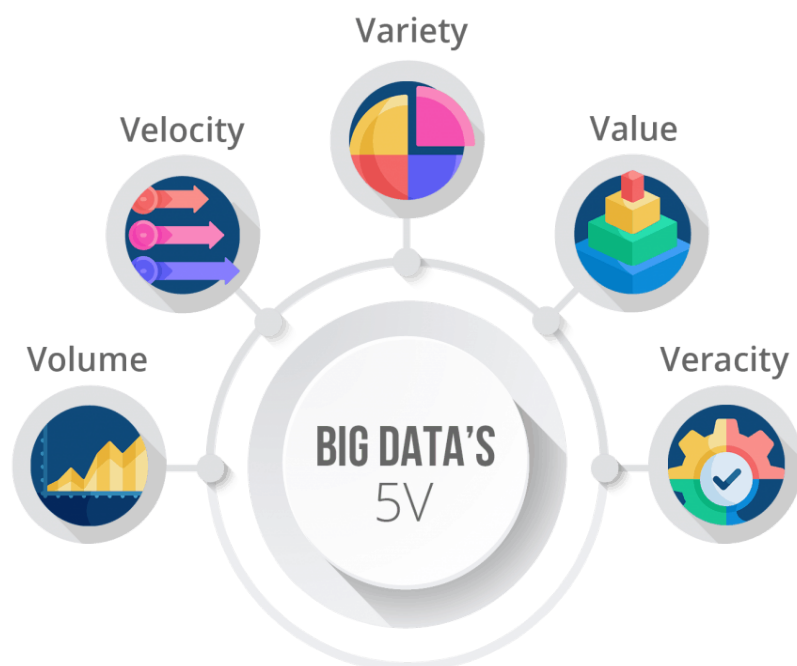
Com a explosão dos dados, surgiram outras questões desafiadoras no tocante à otimização de maneira eficaz da grande massa de dados, visando identificar mecanismos de análise. Somando-se a isso o desafio de transformar essa grande massa de dados em informação e consequentemente em vantagem competitiva para as organizações (RODRIGUES, 2017).

A computação em nuvem, está diretamente relacionada ao *Big Data*, sendo o principal objetivo da computação em nuvem, usar enormes recursos de computação e armazenamento sob o gerenciamento concentrado. O seu desenvolvimento, oferece soluções para o armazenamento e processamento de *Big Data*, acelerando assim seu desenvolvimento. Podendo efetivamente, gerenciar a *Big Data* (CHEN; MAO; LIU, 2014).

### 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). Conforme podemos visualizar na Figura 1.

**Figura 1** Os 5Vs do *Big Data*



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.

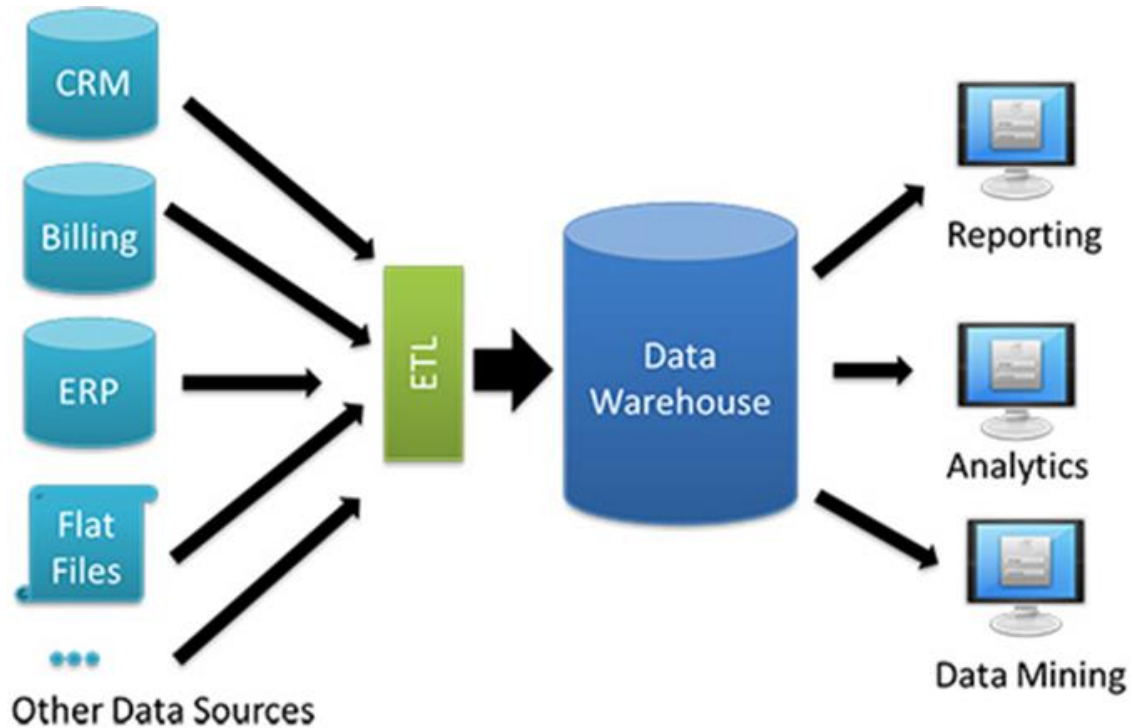
## 2.2 DATA WAREHOUSE

*Data Warehouse* é um repositório que armazena dados estruturados e semiestruturados para relatórios e análises. O *Data Warehouse* pode armazenar grandes quantidades de informações e ajuda as empresas na tomada de decisão (WIJAYA; PUDJOATMODJO, 2015). Conforme a Figura 2, podemos visualizar o fluxo de entrada de dados, o armazenamento desses dados em um *Data Warehouse* e o sua saída.

O *Data Warehouse*, pode ser entendido como uma estratégia de organização de dados com o foco para a gestão estratégica e não para controle do negócio. Os *Data Warehouse*, possuem dados sumarizados, agregados e armazenados por longos períodos de tempo. Sendo uma forma de gerir volumes grandes de dados que

se encontram espalhados em diversos sistemas de uma organização (RASLAN; CALAZANS, 2014).

**Figura 2** Funcionamento do *Data Warehouse*



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

### 2.2.1 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 ETL, 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.

### 2.2.2 PRINCIPAIS CARACTERÍSTICAS DE UM *DATA WAREHOUSE*

Dentre as principais características de um *Data Warehouse*, encontram-se a gestão de dados, que é uma das principais características. O *Warehouse* viabiliza uma visualização histórica dos dados, ao invés de somente uma visualização do estado atual. Tornando a compreensão dos dados ainda mais profunda. Outro fator é a integração dos dados. No *Data Warehouse*, há uma concentração de dados de diferentes fontes, com formatos distintos. Essa integração é a forma fundamental para que os dados possam ser mais bem utilizados, facilitando a sua gestão. Essa centralização torna a estrutura mais complexa, simplificando assim os passos seguintes a partir do *Data Warehouse*. Outra principal característica é a não volatilidade dos dados. Quando os dados se encontram no *Data Warehouse*, eles permanecem persistentes e estáveis, não modificando o seu estado. Mesmo com a capacidade do *Data Warehouse*, visualizar os dados mais antigos em relação aos mais recentes, esses dados permanecem inalterados. Assim como os bancos de dados que compõe a sua estrutura o *Data Warehouse* é fundamentalmente relacional. Fazendo com que os dados não relacionais precisem passar por um tratamento através de ferramentas de ETL antes de serem carregados (SARAVANAMUTHU; NAWAZ, 2015).

### 2.2.3 BENEFÍCIOS DO *DATA WAREHOUSE*

Dentre os benefícios de um *Data Warehouse*, pode-se citar, agilidade nas consultas. Pelo fato de organizar os dados em um formato relacional, de fácil entendimento, os *Warehouses* favorecem consultas e análises rápidas, obtendo assim um melhor desempenho. Outro ponto são os dados de qualidade. A arquitetura do *Data Warehouse*, pressupõe a existência de ferramentas de tratamento dos dados, o que garante que apenas dados tratados e limpos sejam carregados e fiquem disponíveis para o acesso. Segurança, é possível definir regras de acesso, definir regras de acesso, manter o controle do ciclo de vida dos dados enquanto eles estiverem armazenados no *Data Warehouse*. A visão histórica também é um

benefício, pois o *Data Warehouse* não armazenará somente os dados recentes, ele irá gerenciar a mudança com o tempo, é possível analisar tendências temporais e analisar como os dados mudaram e o padrão de mudança desses dados. Escalabilidade, é mais uma característica vantajosa do *Data Warehouse*, demonstrando como a estrutura é a mais indicada para tratar com conjuntos crescentes de dados, gerados em tempo real. Com a solução de tratamento o *Data Warehouse*, conseguirá tratar e organizar os dados, independente do seu tamanho ou do tamanho dos conjuntos em geral. A autonomia para os funcionários pode ser mais um benefício a ser citado, pois a vantagem de um *Data Warehouse*, é desenvolver autonomia para que os dados sejam acessados e analisados para todos os colaboradores, devido a sua estrutura que é organizada visando a facilidade na separação de dados e obtenção de relatórios e consultas (SARAVANAMUTHU; NAWAZ, 2015).

### 2.3 DATA LAKE

*Data Lake* é um repositório que armazena todos os dados estruturados, não estruturados e semiestruturados, com diferentes formatos, permitindo que seja gerado um conjunto de estratégias de negócios. O *Data Lake*, pode armazenar os dados dos 3 tipos, estruturados, semiestruturados e não estruturados, os dados estruturados, são dados que são organizados em uma estrutura predefinida e são de fácil consulta através de consultas SQL, tornando mais fácil a consulta e análise dos dados. Os dados semiestruturados, são uma combinação de dados estruturados e não estruturados, eles não seguem uma estrutura rígida, possuindo um formato mais adaptativo, seguindo geralmente os formatos de armazenamento como XML, JSON e YAML. Os dados não estruturados, são informações que não seguem uma estrutura predefinida, não estando organizados em tabelas ou banco de dados relacionais, podendo incluir arquivos de texto, imagens, áudio e vídeo. Como exemplo desse tipo de dados, pode-se citar, e-mail, documentos, páginas web, *feeds* de mídia social, além de arquivos de imagem e áudio (COLABORATIVO, 2022).

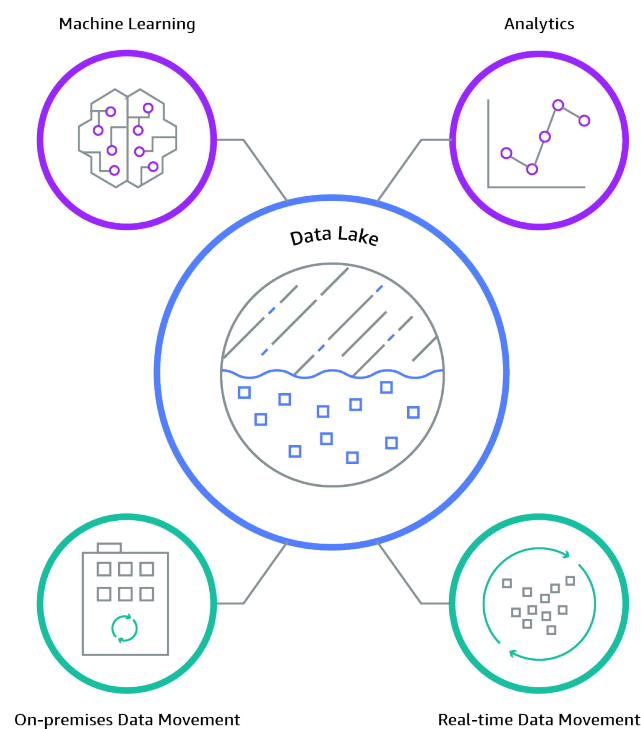
Um *Data Lake*, propõe um armazenamento de dados em seu formato nativo, em local de armazenamento evolutivo. A ideia básica é que todos os dados emitidos pela organização, sejam armazenados em uma única estrutura de dados. Esses dados serão armazenados em seu formato original, dispensando a etapa de

transformação, agilizando a disponibilidade desses dados. Propondo assim um novo modelo de processamento conhecido como *ELT* (RAU, 2021).

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).

Conforme a Figura 3, podemos visualizar a centralização de dados proposta pelo *Data Lake*.

**Figura 3** Funcionamento do *Data Lake*



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:

- 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).

- Armazenamento e catalogação 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).

- 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).

- *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 FASES ESTRATÉGICAS

Foram definidas oito fases na implantação de um *Data Lake*. Fase das fontes de dados, que consiste em selecionar as fontes que serão relevantes para o projeto em questão, essas fontes podem incluir dados de variadas fontes de formatos. Fase de extração e carregamento de dados, que irá consistir na coleta dos dados de entrada armazenados no *Data Lake*, para que sejam implementadas técnicas de ciências de dados com base nos requisitos do projeto. Fase de recuperação da informação, responsável pela implementação de estratégias de busca nos conjuntos de dados extraídos e carregados no *Data Lake*. Fase de tratamento de dados, onde será

executada a aplicação das técnicas de estatística e informática que realizará a análise dos dados recolhidos nas fases anteriores. Fase de construção do modelo, onde serão escolhidos os modelos de aprendizagem computacional ou modelos estatísticos. Fase de avaliação e validação dos resultados, avaliação e validação dos resultados obtidos, serão avaliadas com base nas métricas de desempenho estabelecidas, bem como em processos estatísticos baseados nas melhores práticas. Fase de apresentação dos resultados, apresentação dos resultados e demonstração dos protótipos de produtos de dados criados. Fase de entrega do produto de dados, fase final que irá ser responsável pela implementação dos protótipos aprovados (COLABORATIVO, 2022).

### 2.3.3 ARQUITETURA DO *DATA LAKE*

A arquitetura funcional de um *Data Lake*, consiste em 4 zonas básicas. Zona de dados brutos, todos os dados são armazenados em seu estado nativo, sem serem processados, podendo serem processados em lote ou em tempo real, essa zona é utilizada para encontrar a primeira forma dos dados para análise, auxiliando nos próximos tratamentos. Zona de processo, é a zona onde os dados capturados podem ser alterados de acordo com os requisitos, permitindo que sejam executados os processos de transformação de dados. Zona de acesso, onde os dados serão armazenados e acessados para a realização das análises devidas. Zona de governança, toda as zonas dependem da governança de dados, ela que será diretamente responsável pela segurança dos dados, qualidade e ciclo de vida das informações (SINGH; SINGH; BHATI, 2022).

### 2.3.4 COMPARAÇÃO ENTRE *DATA WAREHOUSE* E *DATA LAKE*

O Quadro 1 ilustra as principais diferenças entre as duas abordagens:

**Quadro 1** *Data Warehouse x Data Lake*

<b>Comparação</b>	<b>Data Warehouse</b>	<b>Data Lake</b>
Dados	Estruturados, dados processados	Estruturados/Semi-estruturados, dados não tratados, dados não processados
Processamento	Esquema na gravação	Esquema na leitura
Armazenamento	Caro, confiável	Baixo custo
Agilidade	Menos ágil, configuração fixa	Alta agilidade, configuração flexível
Segurança	Maduro	Amadurecendo
Usuários	Profissionais de negócio	Cientista de dados (especialistas)

Fonte: Rau (2021). Acesso em: 07/06/2023.

## 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 (RAJ; BOSCH; OLSSON; WANG, 2020).

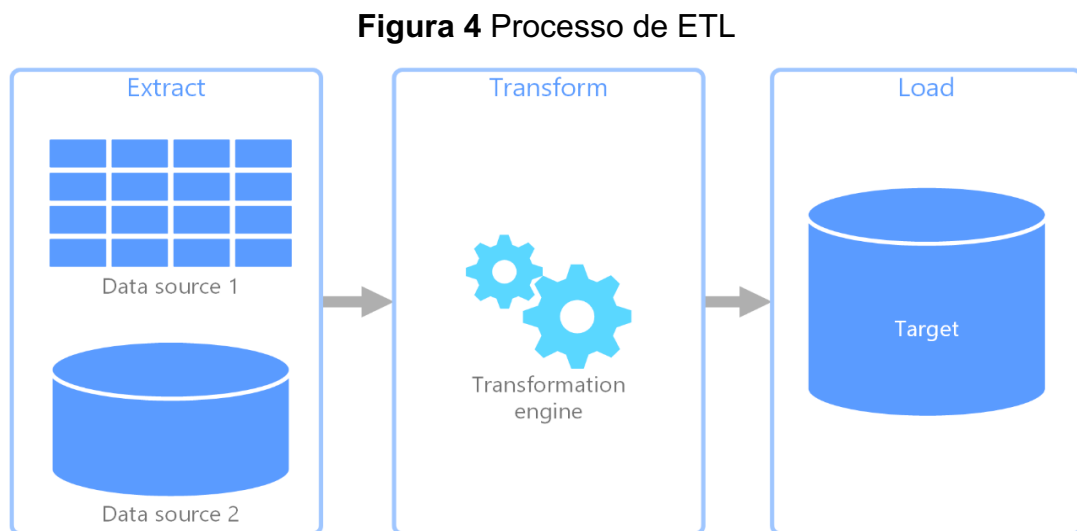
O ETL, é conhecido como sendo um subprocesso da construção de um *Data Warehouse*, pois encontra-se associado aos projetos. O que não significa que ele poderá ser utilizado para outros tipos de sistemas de armazenamento de dados. O ETL tem que ser capaz de suportar dados de diferentes tipos, comunicar-se com base de dados distintas e ler formatos de arquivos diferentes (TAVARES, 2013).

O Processo de ETL, é o mais crítico e complexo, pois irá consumir mais tempo em um projeto de *Data Warehouse*, podendo consumir até 70% dos recursos necessários para implementação. Pois todo o processo desde a coleta de dados,

transformação até o carregamento, é um processo duradouro, complexo e crítico, sendo a transformação a etapa que consome mais tempo. Como exemplo de ferramentas de ETL, pode-se citar, *Oracle Data Integrator*, *Pentaho*, *Apache Sark* e o *SQL Server Integration Services (SSIS)* (TAVARES, 2013).

#### 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 (TIWARI, 2016). Conforme a Figura 4.



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; TIWARI, 2016).

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; TIWARI, 2016):

- Transformação de dados básica
  - Limpeza de dados
  - Eliminação de duplicação de dados
  - Revisão de formato de dados
- Transformação de dados avançada
  - 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; TIWARI, 2016).

O carregamento de dados tem dois métodos:

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

## 2.5 ELT

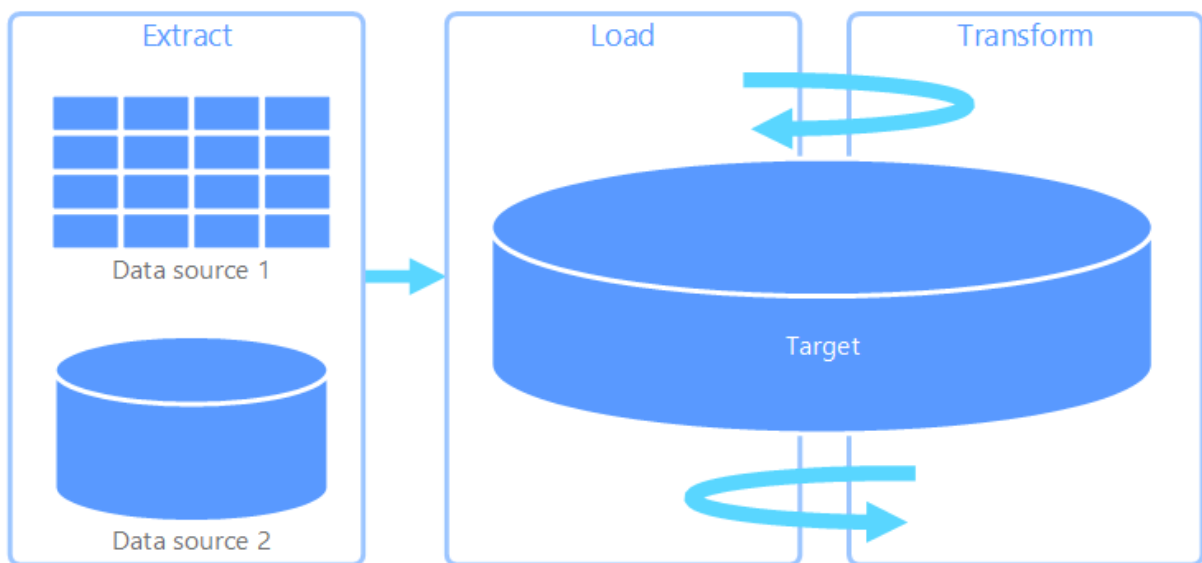
É um pipeline que a transformação ocorre no armazenamento de dados de destino, sem ter a necessidade de usar um mecanismo de transformação separado. Os recursos de processamento do armazenamento de dados de destino são usados para transformar os dados (ZVONAREV; GUDILIN; LYCHAGIN; GORYACHKIN, 2023). Conforme Figura 5, abaixo.



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).

Diferente do ETL, no ELT, os dados serão separados e logo serão empilhados, combinando todas as fontes de dados em um único arquivo integrado. Após a separação e o carregamento dos dados, eles serão apenas uma duplicata da fonte de origem. Com o avanço na estrutura baseada em nuvem, as estruturas agora seriam capazes de suportar a grande capacidade e números flexíveis. Para que se consiga manter o controle de todas as informações brutas extraídas, se faz necessário a utilização de um conjunto de dados grandes e expansivo, bem como o processamento necessário para realização das transformações necessárias. Por isso a arquitetura em nuvem é fator determinante para utilização do ELT (SINGHAL; AGGARWAL, 2022).

**Figura 5** Processo de ELT



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. Assim como no ETL o ELT possui três fases de execução. Extração, que faz a captura de dados brutos de diversas fontes para a integração em um repositório de dados único. O carregamento dos dados coletados em um *Data Warehouse* ou repositório

de dados. A transformação dos dados brutos em dados modelados dentro de um *Data Warehouse* para a aplicação de *Business Intelligence*, análise de dados e *advanced analytics* (SINGHAL; AGGARWAL, 2022).

### 2.5.2 TRANSFORMAÇÃO DE DADOS

Os dados necessitam passar por uma transformação antes que sejam disponibilizados para o usuário final. Essa é a etapa mais difícil dos processos ETL e ELT. O ELT é foi pensado para a utilização em conjunto com um armazenamento em nuvem, pois é possível realizar o armazenamento de dados não relacionais, como banco de dados NoSQL como o MongoDB ou sistema de arquivos Hadoop, Snowflake e o Amazon S3. Esses sistemas são utilizados para ingerir fontes de dados altamente dinâmicas e sem esquema. Para a carga de trabalho analítica, os dados são transformados e processados em formato tabular. As decisões de negócios baseadas em dados, devem ser tomadas em tempo hábil e essa frequência pode variar, semanalmente ou diariamente, para que esses relatórios sejam gerados. Por isso é esperado que esse grande volume de dados seja disponibilizado ao vivo, basicamente como um *streaming* de dados, como exemplo desse tipo de transmissão, pode-se citar a transmissão de dados do mercado de ação. Nessas circunstâncias os aplicativos de ETL, podem ter um funil de processamento em seu processo de transformação, pois a fila de processamento pode transbordar, fazendo com que o usuário espere por um período fora do normal. Para isso foi criado a abordagem ELT, ela foi pensada para resolver esse problema. A localização da transformação de dados é a principal distinção entre ETL e ELT. O ELT ao contrário do ETL, não modifica as informações durante o trânsito, o banco de dados que irá manipular a transformação. Isso significa que os dados serão enviados diretamente para o sistema de armazenamento e por sua vez com os dados dentro do sistema de armazenamento, a lógica de negócios irá entrar em ação e preparando os dados e copiando-os para uma área separada para disponibilização. Todo o gerenciamento é feito pelo provedor em nuvem, desde o dimensionamento do sistema, até o gerenciamento do armazenamento. Como todo processamento é realizado em nuvem, esses sistemas de armazenamento, possuem máquinas com alto processamento em RAM, além de vários núcleos de processadores e armazenamento em disco rápido, mesmo se ocorrer alguma falha de funcionamento os *clusters*, permanecem online. Como exemplo de sistemas de

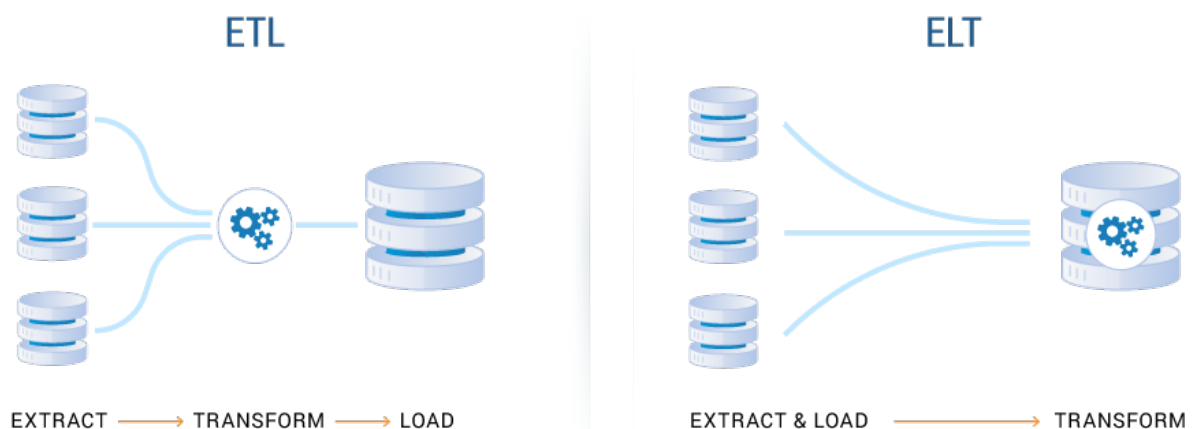
armazenamento em nuvem existem o Amazon *Redshift* e o *Snowflake* (SINGHAL; AGGARWAL, 2022).

### 3 COMPARATIVO ENTRE ETL E ELT

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. Conforme observado na Figura 6. 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 (SINGHAL; AGGARWAL, 2022).

Cada abordagem oferece vantagens e desvantagens a serem consideradas.

**Figura 6 ETL x ELT**



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

### 3.1 VANTAGENS ETL

O ETL equilibra a capacidade e compartilhar o trabalho com o sistema de gerenciamento de banco de dados relacional. Ao utilizar mapas de dados, o ETL executa operações mais complexas em um único diagrama de fluxo de dados. Podendo lidar com segregação e paralelismo. Processamento em lote dos dados transmitidos da origem e carregando no destino. Preserva as plataformas de fonte de dados atuais sem se preocupar com a sincronização de dados. 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. O ETL transfere apenas dados que já foram transformados, com isso pode haver uma economia nos custos de armazenamento armazenando apenas os dados necessários. Suporta gerenciamento de dados complexos, ajudando com cálculos, integrações de dados e manipulação de *strings*. Segurança e conformidade na criptografia de dados confidenciais (FÁTIMA, 2020; IBM, 2021; RAJ, BOSCH, OLSSON, WANG, 2020; TAVARES, 2013; TIWARI, 2016; SINGHAL, AGGARWAL, 2022).

### 3.2 DESVANTAGENS ETL

O ETL necessitará de despesas adicionais de *hardware*. O processamento das informações linha a linha poderá diminuir o desempenho do processo de ETL. Flexibilidade reduzida devido à dependência de fornecedores de ferramentas. Os dados necessitam ser transferidos para uma camada adicional antes de chegarem ao destino. Não existem controle programado de erros ou mecanismo de recuperação. O processo necessita de atualizações periódicas ao invés de atualizações em tempo real. Tempo mais longo de carregamento devido ao grande número de etapas no estágio de transformação. Maior tempo de desenvolvimento (FÁTIMA, 2020; IBM, 2021; RAJ, BOSCH, OLSSON, WANG, 2020; TAVARES, 2013; TIWARI, 2016; SINGHAL, AGGARWAL, 2022).

### 3.3 VANTAGENS ELT

Melhor desempenho e segurança de dados. Necessita de menos tempo e recursos, pois os dados serão transformados e carregados em paralelo. Permite dados de tamanho maiores. Não necessita de um bloco de transformação, pois o

sistema de destino que realiza esse trabalho. Maior flexibilidade, não exigindo o desenvolvimento de pipelines complexos. A transformação de dados é realizada apenas para os dados necessários para que seja realizada uma análise específica. Implementação mais rápida. Utilização de sistemas de armazenamento em nuvem. Suporte a tipos de dados estruturados, não estruturados, semiestruturados e brutos (FÁTIMA, 2020; IBM, 2021; RAJ, BOSCH, OLSSON, WANG, 2020; TAVARES, 2013; TIWARI, 2016; SINGHAL, AGGARWAL, 2022).

### 3.4 DESVANTAGENS ELT

Limitação nas ferramentas que oferecem suporte para o ELT. Falta de modularidade devido ao design baseado em conjunto. O carregamento de dados confidenciais antes de transformá-los, expõe os dados privados em logs acessíveis aos administradores do sistema. A evolução recente causa uma falta de confiança nas ferramentas de ELT em relação as de ETL. Demanda uma maior quantidade de armazenamento (FÁTIMA, 2020; IBM, 2021; RAJ, BOSCH, OLSSON, WANG, 2020; TAVARES, 2013; TIWARI, 2016; SINGHAL, AGGARWAL, 2022).

### 3.5 PRINCIPAIS DIFERENÇAS

O Quadro 2 ilustra o comparativo entre os modelos de extração ETL e ELT:

**Quadro 2 ETL x ELT**

	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: Singhal e Aggarwal, (2022), acesso em 10/06/2023.

#### 4 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).

Analisando o funcionamento de ambos os processos, pode-se tirar a conclusão de que para sistemas monolíticos, cujo dados não possuam uma alta variação em sua origem e que possuam uma infraestrutura convencional de data center, o ETL será o processo mais indicado a ser implantado, pois é um processo já consolidado, onde ele atenderá as necessidades. Para arquiteturas baseadas na nuvem, onde os bancos de origem possuem uma variabilidade maior nos dados, podendo ser dados estruturados ou não estruturados, o ELT será o processo mais indicado, pois ele irá atender a demanda de recuperar informações provenientes de qualquer base origem, disponibilizando essa informação de maneira mais rápida, pois utilizará de todo o poder de processamento dos servidores em nuvem. Com o avanço dessa tecnologia o ELT tenderá cada vez mais a ser utilizado, afim de atender a alta dinamicidade que os dados estão alcançando cada vez mais (SINGHAL; AGGARWAL, 2022).

## 5 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. 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 Madiseti. 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.



CHEN, Min; MAO, Shiwen; LIU, Yunhao. Big Data: a survey. **Mobile Networks And Applications**, [S.L.], v. 19, n. 2, p. 171-209, 22 jan. 2014. Springer Science and Business Media LLC. <http://dx.doi.org/10.1007/s11036-013-0489-0>. Acesso em: 07 jun. 2023.

COLABORATIVO, Grupo. Data Lake Strategy for Data Science Workflows. **2022 11Th International Conference On Software Process Improvement (Cimps)**, [S.L.], v. 2022, n. 11, p. 101-105, 19 out. 2022. IEEE. <http://dx.doi.org/10.1109/cimps57786.2022.10035686>. Disponível em: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10035686&isnumber=10035565>. Acesso em: 09 jun. 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.

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\\_9EkvWfp2MfeKGGGo6tQZZ8Vq7O0HjCcs0lZVzuOso-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_9EkvWfp2MfeKGGGo6tQZZ8Vq7O0HjCcs0lZVzuOso-zlaAirnEALw_wcB). Acesso em: 14 mar. 2023.

RAJ, Aiswarya; BOSCH, Jan; OLSSON, Helena Holmstrom; WANG, Tian J.. Modelling Data Pipelines. **2020 46Th Euromicro Conference On Software Engineering And Advanced Applications (Seaa)**, [S.L.], v. 46, n. 2020, p. 13-20, ago. 2020. IEEE. <http://dx.doi.org/10.1109/seaa51224.2020.00014>. Disponível em: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9226314&isnumber=9226268>. Acesso em: 09 jun. 2023.

RASLAN, Daniela Andrade; CALAZANS, Angélica Toffano Seidel. Data Warehouse: conceitos e aplicações - doi. **Universitas: Gestão e TI**, [S.L.], v. 4, n. 1, p. 25-37, 4 ago. 2014. Centro de Ensino Unificado de Brasília. <http://dx.doi.org/10.5102/un.gti.v4i1.2612>. Acesso em: 07 jun. 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.

RODRIGUES, A. A.; DUARTE, E. N.; DIAS, G. A. Desafios da gestão de dados na era do big data: perspectivas profissionais \*. **Encontro Nacional de Pesquisa e Pós-graduação em Ciência da Informação**, n. XVIII ENANCIB, 2017. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/104921>. Acesso em: 07 jun. 2023.

SARAVANAMUTHU, M.; NAWAZ, G. M. Kadhar. Maximum performance with minimum cost in data mining applications through the novel online data warehouse architecture by using storage area network with fibre channel fabric. **2015 International Conference On Circuits, Power And Computing Technologies [Iccpct-2015]**, [S.L.], v. 2015, n. 2015, p. 1-7, mar. 2015. IEEE. <http://dx.doi.org/10.1109/iccpct.2015.7159498>. Disponível em: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7159498&isnumber=7159156>. Acesso em: 08 jun. 2023.

SEGNER, Michael. **ETL vs ELT: What's the Difference (and Which is Better)?** 2023. Disponível em: <https://www.montecarloata.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.

SINGH, Jaspreet; SINGH, Gurpreet; BHATI, Bhoopesh Singh. The Implication of Data Lake in Enterprises: a deeper analytics. **2022 8Th International Conference On Advanced Computing And Communication Systems (Icaccs)**, [S.L.], v. 8, n. 2022, p. 530-534, 25 mar. 2022. IEEE. <http://dx.doi.org/10.1109/icaccs54159.2022.9784986>. Disponível em: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9784986&isnumber=9784957>. Acesso em: 09 jun. 2023.

SINGHAL, Bharat; AGGARWAL, Alok. ETL, ELT and Reverse ETL: a business case study. **2022 Second International Conference On Advanced Technologies In Intelligent Control, Environment, Computing & Communication Engineering (Icatiece)**, [S.L.], v. 2, n. 2022, p. 1-4, 16 dez. 2022. IEEE. <http://dx.doi.org/10.1109/icatiece56365.2022.10046997>. Disponível em: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10046997&isnumber=10046669>. Acesso em: 09 jun. 2023.

TAVARES, Edmir de Jesus Oliveira. **O Processo ETL: o caso da unitel t + telecomunicações**. 2013. 133 f. TCC (Graduação) - Curso de Engenharia de Sistemas e Informática, Universidade Jean Piaget de Cabo Verde, Cidade da Praia, 2013. Disponível em: <https://core.ac.uk/download/pdf/38682758.pdf>. Acesso em: 07 jun. 2023.

TIWARI, Prayag. Advanced ETL (AETL) by integration of PERL and scripting method. **2016 International Conference On Inventive Computation Technologies (Icict)**, [S.L.], v. 2016, n. 2016, p. 1-5, ago. 2016. IEEE. <http://dx.doi.org/10.1109/inventive.2016.7830102>. Disponível em: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7830102&isnumber=7830057>. Acesso em: 09 jun. 2023.

WIJAYA, Rahmadi; PUDJOATMODJO, Bambang. An overview and implementation of extraction-transformation-loading (ETL) process in data warehouse (Case study: department of agriculture). **2015 3Rd International Conference On Information And Communication Technology (Icoict)**, [S.L.], v. 3, n. 2015, p. 70-74, maio 2015. IEEE. <http://dx.doi.org/10.1109/icoict.2015.7231399>. Disponível em: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7231399&isnumber=7231384>. Acesso em: 08 jun. 2023.

ZVONAREV, Aleksei E.; GUDILIN, Dmitriy S.; LYCHAGIN, Dmitriy A.; GORYACHKIN, Boris S.. Extract-Load-Transform (ELT) Process Runtime Analysis and Optimization. **2023 5Th International Youth Conference On Radio Electronics, Electrical And Power Engineering (Reepe)**, [S.L.], v. 5, n. 2023, p. 1-7, 16 mar. 2023. IEEE. <http://dx.doi.org/10.1109/reepe57272.2023.10086728>. Disponível em: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10086728&isnumber=10086686>. Acesso em: 09 jun. 2023.